

JRC TECHNICAL REPORTS

Inter-sensor comparison of built-up derived from Landsat, Sentinel-1, Sentinel-2 and SPOT5/SPOT6 over selected cities

Filip Sabo, Christina Corbane, Stefano Ferri

2017



This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

Contact information

Name: Christina Corbane

Address: European Commission, Joint Research Centre, Space, Security and Migration (Ispra), Disaster Risk Management (JRC.E.1)

E-mail: christina.corban@ec.europa.eu

Tel.: +39 0332 78 3545

JRC Science Hub

<https://ec.europa.eu/jrc>

JRC105855

EUR 28520 EN

PDF	ISBN 978-92-79-66706-0	ISSN 1831-9424	doi:10.2760/385820
Print	ISBN 978-92-79-66705-3	ISSN 1018-5593	doi:10.2760/ 71297

Luxembourg: Publications Office of the European Union, 2017

© European Union, 2017

The reuse of the document is authorised, provided the source is acknowledged and the original meaning or message of the texts are not distorted. The European Commission shall not be held liable for any consequences stemming from the reuse.

How to cite: Sabo F., Corbane C., Ferri S. Inter-sensor comparison of built-up derived from Landsat, Sentinel-1, Sentinel-2 and SPOT5/SPOT6 over selected cities. EUR 28520 EN, doi:10.2760/385820

All images © European Union 2017

Table of contents

Abstract	1
1. Introduction	1
2. Overview of the four built-up products and their classification methods	3
2.1 The SML classifier	3
2.1.1 Built-up extraction from Landsat (GHSL-Landsat)	4
2.1.2 Built-up extraction from Sentinel-1	5
2.1.3 Built-up extraction from Sentinel-2	6
2.2 European Settlement Map based on SPOT-5 and SPOT-6	7
3. Study sites and reference data	8
4. Validation methodology	9
4.1 Performance metrics	10
4.2 Built-up density analysis	11
4.3 Regression analysis of built-up	12
5. Results and discussion	12
5.1 Results of performance metrics	12
5.2 Results of performance metrics	14
5.3 Results of regression analysis of built-up	17
6. Visual inspection	23
7. Conclusion	32
Acknowledgements	33
Annex	34
References	36
List of abbreviations and definitions	38
List of figures	39
List of tables	41

Abstract

In the last 5 years, several information layers describing human settlements were developed within the Global Human Settlement infrastructure of the Joint Research Centre using Earth Observation data. Each layer was derived from a different satellite (with different various spatial resolutions and radiometric properties) and from images acquired at different time stamps. The next step is to exploit the synergies between the different sensors and possibly integrate the information layers within a single product. To enable those future developments, it is essential to understand the potentials and limitations of each of the layers and identify complementarities. In these regards, the validation of built-up derived from different sensors is crucial for gaining a deeper understanding of the consistency and interoperability between them. This report, presents the methodology and the results of the inter-sensor comparison of built-up derived from Landsat, Sentinel-1, Sentinel-2 and SPOT5/SPOT6.

The assessment was performed for 13 cities across the world for which fine scale reference building footprints were available. Several validation approaches were used: cumulative built-up curve analysis, pixel by pixel performance metrics and regression analysis. The results indicate that the Sentinel-1 and Sentinel-2 highly contribute to the improved built-up detection compared to Landsat. However, Sentinel-2 tends to show high omission errors while Landsat tends to have the lowest omission error. The built-up obtained from SPOT5/SPOT6 show high consistency with the reference data for all European cities and hence can be potentially considered as a reference dataset for wall-to-wall validation at the European level. It was noted that the validation results can highly vary across all the study sites because of the different landscapes, settlement structures and densities.

1. Introduction

Global information about human presence is crucial for different application areas such as risk assessment, infrastructure planning, urban expansion, poverty reduction, biodiversity conservation and others (Chrysoulakis et al., 2014; Florczyk et al., 2016; Triantakou et al., 2015). Considering the availability of open and free earth observation data (Copernicus Sentinel-1 and Sentinel-2 data, Landsat imagery), information about global human presence on Earth can be precisely and timely updated and refined. Mapping and monitoring human settlements using satellites and automatic information extraction methods on a global level is a challenging issue. This is because built-up is composed of different roof materials, present different shapes and colours, and are characterized by a high heterogeneity in the signal recorded by the satellite sensor.

Several initiatives attempted to map human settlements at the global level. Esch et al. (2013) developed the Global Urban Footprint processor which is a fully automated approach for mapping human settlements using commercial Synthetic Aperture Radar data. A total of 180 000 TerraSAR-x and TanDEM-x scenes were processed in order to map the built-up across the globe with a spatial resolution of 12 m.

Other studies that focused on global land cover mapping by using Landsat data are reported in (Gong et al., 2013): the Finer Resolution Observation and Monitoring of Global Land Cover (FROM-GLC) and GlobeLand30. However, these studies failed to map the urban areas properly in terms of accuracy (Gong et al., 2013) or are difficult to reproduce and update due to the high level of human intervention in the production of the layers.

Recently, Ban et al. (2015) used ENVISAT SAR data for developing robust urban extractor across 10 selected cities. The validation of the urban extractor produced good overall accuracy of 85.4 %, commission error of 23.7% and omission was 5.47%, averaged for 10 cities.

The Global Human Settlement Layer (GHSL) was introduced by the Joint Research Centre (JRC) in the years 2010-2011 as a framework (Pesaresi et al., 2011) aiming to provide improved, ready-to-use or pre-calculated baseline data reporting about the human presence on the globe. It includes new and open tools and methods for automatic information extraction of built-up from remote sensing data. In the framework of the GHSL, several layers describing human settlements at the European and Global levels were produced from satellite imagery and other future products are currently under development. In 2014, an innovative solution for seamless continental-wide, systematic information extraction and mapping using SPOT5/6, 2.5 m resolution input data was developed with the GHSL tools. The final outcome was the European Settlement Map which represents currently the widest area ever mapped by applying automatic data classification techniques (Ferri et al., 2014, 2017; Florczyk et al., 2016).

In 2015, in the framework of the GHSL, the first world-complete human settlements data was generated from historical records of Landsat imagery organized in four collections (1975, 1990, 2000 and 2014) at a spatial resolution of 30 meters. A novel data mining and machine learning technology, the Symbolic Machine Learning (SML) classifier (M. Pesaresi et al., 2016b) was deployed for the fully automated built-up information extraction from Landsat data. Recently, the same technology was adapted to the processing of radar data and successfully applied to a global coverage of Copernicus Sentinel-1 data¹. The output of this experiment was a new global built-up layer at a spatial resolution of 20m. In view of the future update of the GHSL layer at the global scale with Copernicus Sentinel-2 data, the SML has been adapted to the classification of

¹ http://ghsl.jrc.ec.europa.eu/s1_2017.php

Sentinel-2 images. The prototype has been tested on the first Sentinel-2 images released during the commissioning phase (Martino Pesaresi et al., 2016a). The study confirmed the noticeable improvement of Sentinel-2 in comparison to Landsat for built-up classification. The prototype, currently in its final development phase, has been also tested on a set of cities across the world.

Despite the successful adaptation of the SML to the analysis of several types of satellite data, with different spectral and spatial resolutions, there is currently a lack of assessment of the consistency among the derived products and insufficient information on accuracy evaluation. The users of these products should get a minimum guidance on which dataset to be used and for which purpose. Besides, to fully exploit the synergies between the different sensors and possibly integrate those products in the future, it is essential to understand the potentials and limitations of each of them and identify complementarities. In these regards, the validation of built-up derived from different sensors is crucial for gaining a deeper understanding of the consistency and interoperability between them. In the specific case of the Sentinel-2 derived built-up, validation is necessary for improving the workflow currently under development by providing insights into the main issues that need further research and experimentation.

The aim of this report is to provide clear, unbiased validation and comparison of the built-up derived from different input sensors: Spot5/6 in the case of the European Settlement Map, Landsat-8 in the case of the GHSL-Landsat (GHSL-Landsat), Sentinel-1 in the case of a newly released built-up layer and Sentinel-2 for which the workflow for built-up extraction is currently in the prototyping and testing phase. As a benchmark, fine scale building footprints were used as reference data. The analysis was performed over 13 cities selected from different continents, with diverse landscapes and settlement structures and densities. To paint a complete picture, we developed and applied a validation framework that incorporates conventional methods of pixel-by-pixel accuracy assessment and analysis of grid-based differences in built-up in relation to settlements densities. Through this detailed validation, the report attempts to identify main differences in the detection of built-up across sensors and identify potential areas of synergies between the different products.

2. Overview of the four built-up products and their classification methods

We present here a brief description of the four built-up products under assessment together with an overview of the information extraction workflows developed for generating them from different satellite sensors (Landsat, Sentinel-1, Sentinel-2 and SPOT5/SPOT6).

The GHSL built-up areas are defined as spatial units where buildings or part of it can be found (Pesaresi et al., 2013). The four built-up products under analysis share the same working definition of built-up which is defined as follows:

‘buildings are enclosed constructions above ground which are intended or used for the shelter of humans, animals, things or for the production of economic goods and that refer to any structure constructed or erected on its site’. This working definition is adapted from the data specification on buildings delivered by the Infrastructure for Spatial Information in Europe (INSPIRE)², taking in to account the specific GHSL constraints and user requirements. In particular, by contrast to the INSPIRE definition, the GHSL definition does not include underground building notion for obvious limitations of the considered input data (Pesaresi et al., 2013).

The adoption of this definition ensures semantic interoperability across the four products generated in the context of the GHSL despite the different input sensors. It also facilitates direct comparison of the extracted built-up information with reference building footprints. Furthermore, the GHSL classification scheme with its simplification and reduction of the embedded abstraction of the “built-up” concept was designed to facilitate multi-disciplinary across-application sharing of data and results. This includes the sharing of data between different stakeholders working in similar areas, but not necessarily sharing exactly the same abstract definitions (Pesaresi and Ehrlich, 2009).

The technology at the core of the GHSL-Landsat, the Sentinel-1 and Sentinel-2 built-up products relies on the Symbolic Machine Learning (SML) supervised classifier (M. Pesaresi et al., 2016b). The basic concepts of the SML methodology are briefly presented in the next section. The application of the SML to the classification of Landsat data for the generation of the GHSL-Landsat is also recalled and the adaptation of the methodology to the analysis of Sentinel-1 and to Sentinel-2 is concisely introduced.

A brief overview of the methodology used for the generation of the European Settlement Map from SPOT imagery is also given here for informative purposes.

2.1 The SML classifier

The SML is a new generic supervised classification framework developed at the JRC³ which provides a scalable solution to complex and large multiple-scene satellite data processing (M. Pesaresi et al., 2016c).

The SML schema is based on two relatively independent steps:

1. Reduce the data instances to a symbolic representation (unique discrete data-sequences);

² INSPIRE Infrastructure for Spatial Information in Europe, ‘D2.8.III.2 Data Specification on Building – Draft Guidelines’, INSPIRE Thematic Working Group Building 2012 URL:

http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_BU_v2.0.pdf

³ <http://ghsl.jrc.ec.europa.eu/tools.php>

2. Evaluate the association between the unique data-sequences subdivided into two parts: X (input features) and Y (known class abstraction).

The association is measured as a confidence index referred herein as Evidence-based Normalized Differential Index (ENDI) that provides a continuum of positive and negative values ranging from -1 to 1. ENDI expresses strength of association between the image data layers and the reference data. Values close to 1 indicate that the data sequence is strongly associated with the image class of interest – the built-up in our case - while values close to -1 indicate that the feature is strongly associated with the classes other than built-up.

The technology is inspired from DNA microarrays data analysis methods used in biomedical informatics for the clustering of gene expressions. By analogy with the genetic association, the SML classifier developed at the JRC searches for systematic relationships between sequences of satellite data instances (i.e. reflectance values of Landsat or Sentinel-2 ; backscatter intensity of Sentinel-1) and the “roofed built-up” class abstraction encoded in global reference sets.

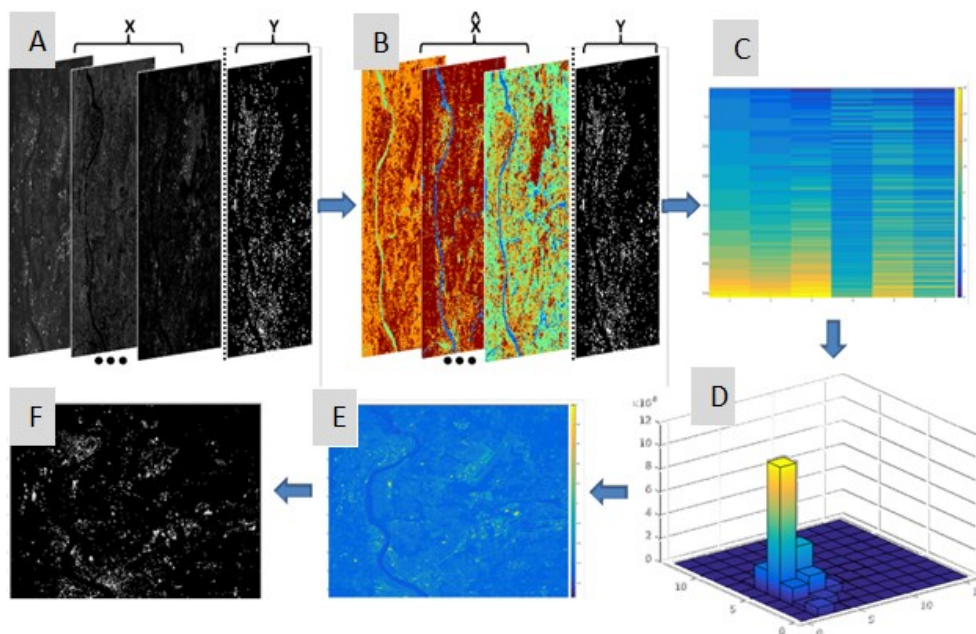


Figure 1. Conceptual processing flow for Symbolic Machine Learning. (A) The image to be classified (X) and the reference training layer (Y); (B) the features , (C) the data sequences, (D) the frequency of association; (E) the confidence index; and (F) the built-up (figure modified from Pesaresi et al., 2016b)

2.1.1 Built-up extraction from Landsat (GHSL-Landsat)

The GHSL-Landsat product was generated with an information extraction technique based on SML. The input data includes four Landsat data collections for 1975, 1990, 2000 and 2014. The classification of Landsat data in the GHSL framework uses reference data for supervised learning composed of coarse scale (i.e. MODIS global urban extents, MERIS GlobCover and LandScan population density grid) and fine scale data (e.g. Open Street Map (OSM), settlement polygons extracted from the urban classes of CORINE and AFRICOVER land cover classes).

The GHSL-Landsat product which was validated in this study corresponds to the epoch 2014, i.e. Landsat images acquired during 2013-2014.

Details on the information extraction procedure are described in (M. Pesaresi et al., 2016a). The following diagram summarizes the information extraction procedure including: the classification of each single epoch, followed by the mosaicking procedure

within a specific epoch, and finally a between-epoch mosaicking procedure in which the information is verified for consistently using logical rules at the multi-temporal level for the production of the final four built-up mosaics.

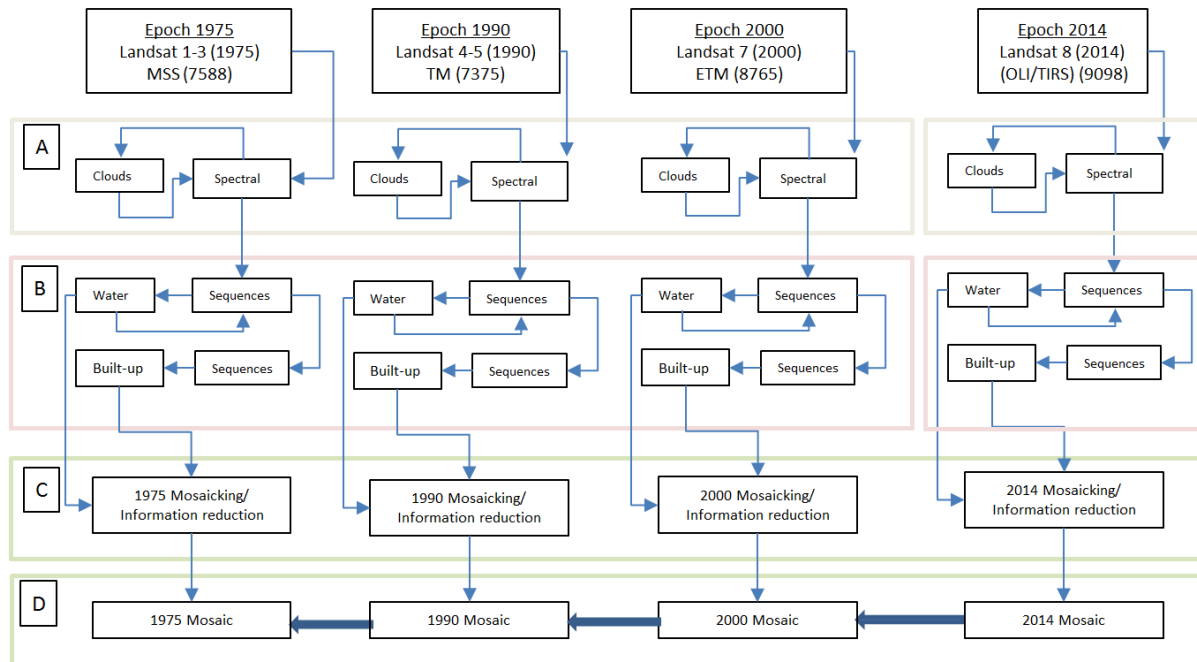


Figure 2. The Landsat GHSL workflow is broken down into four parts. Cloud removal (A), classification of built-up and water (B), data reduction and single date mosaic production (C) and the multi-temporal fusion (D).

2.1.2 Built-up extraction from Sentinel-1

In 2016, the availability of a global coverage of high resolution SAR data collected from the European Sentinel-1 mission was a motivation for testing the applicability of the SML to this new imagery in view of improving and updating the GHSL-Landsat.

The SML workflow was adapted to exploit the key features of the Sentinel-1 Ground Range Detected (GRD) data which are : i) the spatial resolution of 20m with a pixel spacing of 10m and ii) the availability of dual polarisation acquisitions (VV and VH) widely used for monitoring urban areas since different polarizations have different sensitivities and different backscattering coefficients for the same target (Matsuoka and Yamazaki, 2004).

The input features to the SML classifier consisted of: 1) dual polarized backscatter intensities processed at a resolution of 20 m, 2) the mean and standard deviations of backscatter intensities calculated for different windows sizes (3x3, 5x5, 7x7 and 9x9) and 3) topographic features (slope, aspect and crest lines) derived from a digital elevation model (SRTM 90m Digital Elevation Database v4.1)⁴ with the purpose of attenuating the confusion between built-up and topographic structures.

The learning data at the global level consisted of the union of the built-up obtained from the GHSL-Landsat for 2014 and the Global Land Cover map at 30 meter resolution (GLC-30). The latter has been also derived from Landsat imagery through operational visual analysis techniques (Chen et al., 2015).

⁴ <http://www.cgiar-csi.org/data/srtm-90m-digital-elevation-database-v4-1>

A simplified workflow of the adapted SML workflow for the classification of Sentinel-1 (S1) data is shown in Figure 3 with a total of 21 input features and the combined Built-up from the GHSL-Landsat and the GLC30 used for learning in the association analysis.

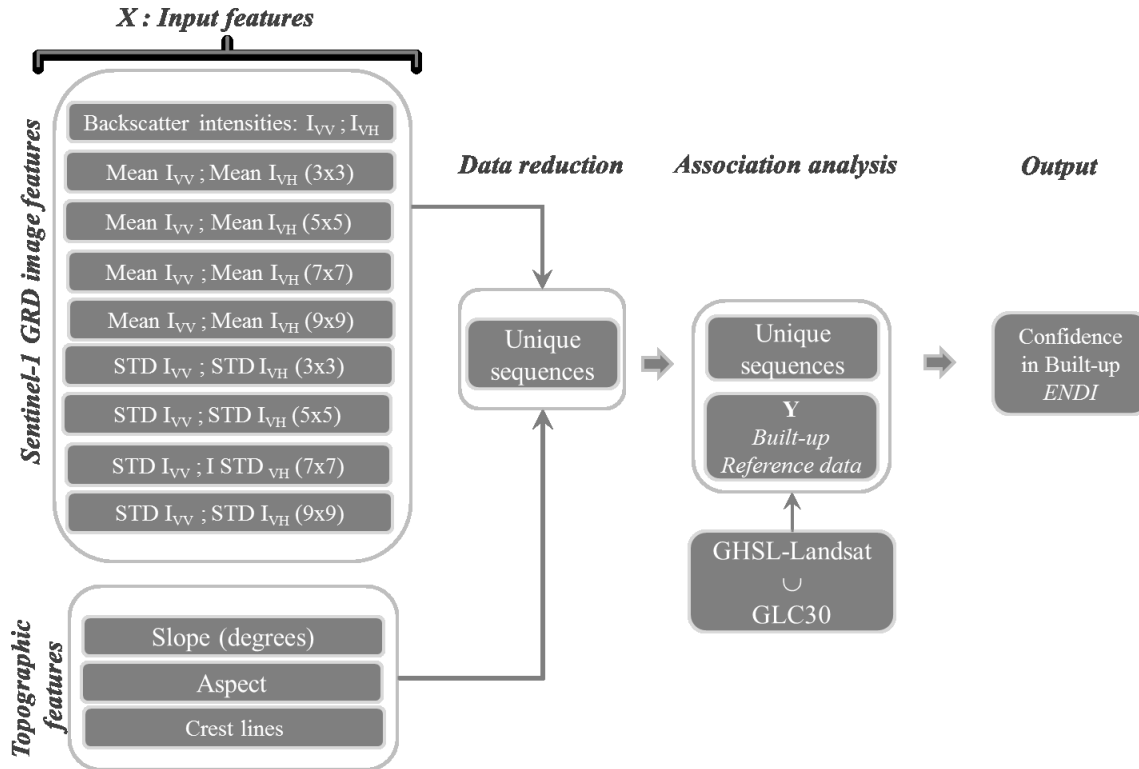


Figure 3. Simplified workflow showing the adaptation of the SML to the classification of Sentinel-1 images at the global level. The input features comprise 18 features derived from dual-polarization Sentinel-1 intensity data and 3 topographic features derived from a global digital elevation model (i.e. SRTM).

2.1.3 Built-up extraction from Sentinel-2

At the time of writing this report, the prototype for the extraction of built-up from Sentinel-2 imagery was still under development. However, an almost stable version of the algorithm was being tested in different landscapes with satisfactory results. The algorithm builds on the SML with adjustments designed for exploiting the key features of Sentinel-2 data: i) the availability of four 10m spatial resolution bands (B2-Blue, B3-Green, B4- Red and B8- Near Infrared), ii) the availability of six bands at 20m resolution especially in the Near Infrared and Shortwave Infrared (B5, B6, B7, B8a in Near Infrared and B11, B12 in Shortwave Infrared).

The following features derived from Sentinel-2 are used for the classification of the Sentinel-2 image with the SML approach:

- Spectral features: the four 10 m resolution and the six 20 m bands
- Textural features: four textural features were derived from the four input 10 m bands by applying the Pantex methodology of (Pesaresi et al., 2008). Those four features were combined in a single feature by using the minimum operator. The textural feature is used for refining the output confidence layer by eliminating overdetections, especially roads and open spaces identified as built-up.

The learning set is based on the built-up as derived from the GHSL-Landsat. The Sentinel-2 images used in the classification are first atmospherically and terrain corrected (Level 2 A outputs are referred to Bottom of Atmosphere BOA) prior to the classification. This is relevant especially when deriving other thematic layers from Sentinel-2 such as the vegetation layer and for monitoring large areas at continental scale, both of which are foreseen for the global GHSL product to be derived from Sentinel-2 during 2017. One of the outputs of the atmospheric correction of Sentinel-2 is scene classification at 20 m resolution with 10 different classes. This rough classification is used in the SML classifier for stratifying the learning set of built-up derived from GHSL-Landsat. This allows tailoring the training set to the image under processing especially in the presence of clouds or cloud shadows and hence allows reducing commission and omission errors.

The following diagram presents a simplified version of the workflow for the classification of Sentinel-2 (S2) data.

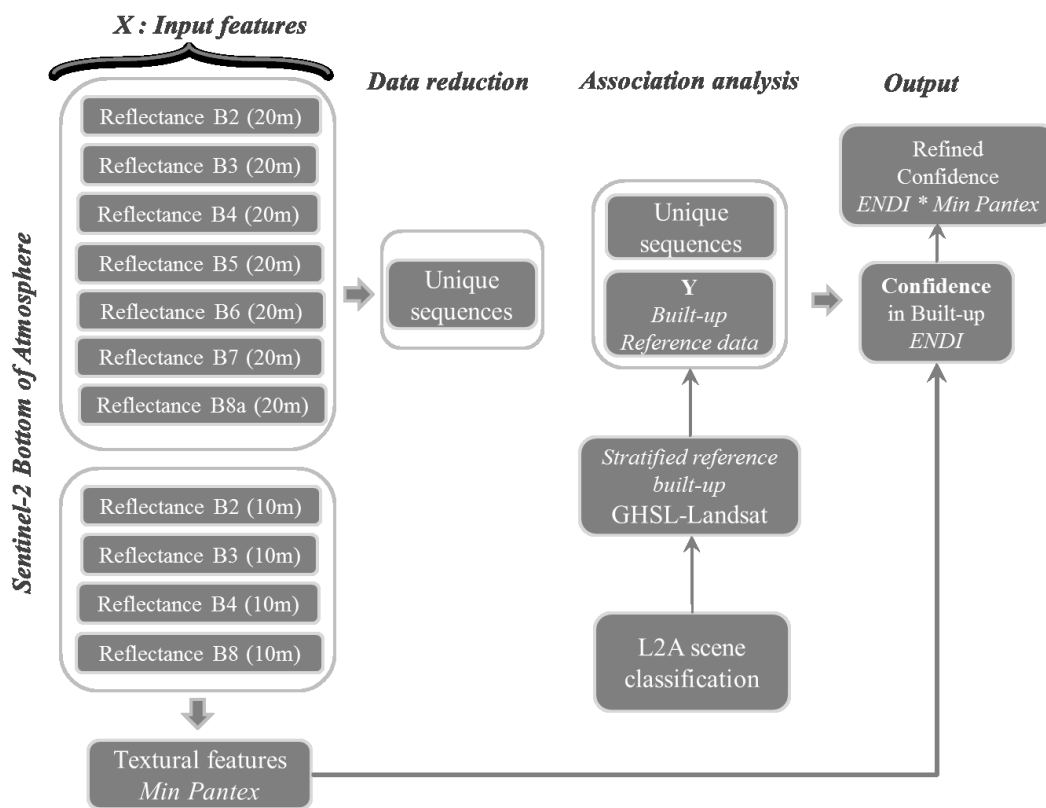


Figure 4. Simplified workflow showing the adaptation of the SML to the classification of built-up from Sentinel-2 BOA images (excluding the extraction of additional thematic layers e.g. vegetation, water, etc.). The input features comprise reflectance values of the four 10 m bands resampled at 20 m and the reflectance values of the 20m bands. The original 10 bands are used for deriving a textural feature that is used for post-processing the ENDI confidence and refining the detection of built-up.

2.2 European Settlement Map based on SPOT-5 and SPOT-6

The European Settlement Map (ESM) is the first high-resolution built-up layer for Europe at 10-m spatial resolution (aggregated from 2.5 m). The ESM classification approach uses more than 3500 satellite images from Spot5 and Spot6 sensors as input data with spatial resolutions of 2.5 m and 1.5 (Florczyk et al., 2016). Total mapped surface amounts to more than 10 million km². The ESM production methodology is strongly based on the GHSL workflow introduced in (M. Pesaresi et al., 2016a). However in the

ESM, the GHSL workflow is modified in order to allow the input data of 2.5 m and it introduces the vegetated surface detector. The main characteristics of the developed methodology are a scene-based processing and a multiscale learning paradigm that combines auxiliary datasets with the extraction of textural and morphological image features. Several auxiliary datasets are used at different stages of the layer production, such as learning, classification, and masking. They include the degree of soil sealing at 100 m (SSL version2), the Corine Land Cover of 2006 also at 100m resolution, the Urban Atlas at a scale of 1:10 000 scale and datasets derived from Open Street Map (buildings, roads, etc.). The validation of the ESM layer against the Land Use/Cover Area frame Statistical Survey (LUCAS) data shows an overall accuracy of 96% with omission and commission errors lower than 4% and 1%, respectively. Full details on the ESM workflow and validation can be found in (Florczyk et al., 2016).

The European Settlement Map which is validated in this report corresponds to the new released ESM with spatial resolution of 2.5 m. (Ferri et al., 2017)

Table 1 provides a summary of the layers analysed in this study. Input spatial resolution corresponds to the pixel size of image input data, whilst the output spatial resolution is the resolution of the final released product. The column year corresponds to the image acquisition periods. For the ESM, Spot 5 and 6 images were acquired during 2010, 2011 and 2013 year. It should be noted that S2 layer is still under production but the results presented here were derived from S2 images acquired in 2016.

Table 1. Characteristics of the built-up products under validation

Product	Spatial resolution [m]		Year	Coverage	Datasource
	Input	Output			
ESM	1.5,2.5	2.5	2010-2013	Europe	http://land.copernicus.eu/pan-european/GHSL
Sentinel-1	10	20	2016	Global	http://ghsl.jrc.ec.europa.eu/ghs_bu_s1.php
Sentinel-2	10,20	20	2016	Selected cities	Under production
GHSL-Landsat	30	38	2013-2014	Global	http://ghsl.jrc.ec.europa.eu/ghs_bu.php

3. Study sites and reference data

The study sites were selected on the basis of availability of reliable reference data and the need to cover different types of built-up structures, settlements densities and landscapes. Reference data consist of fine scale building footprints obtained from different sources: national mapping agencies, national geoportals and Open Street Map (OSM) official. Building footprints downloaded from the OSM were first visually inspected with the help of very high resolution images to ensure completeness of the geographic coverage. A very conservative approach was adopted in which only those areas fully covered by building footprints were considered in the analysis. This approach ensured that gaps in the OSM reference data are avoided. The reference building footprints were first rasterized to 1m pixel size to preserve the maximum possible detail. All the other layers were resampled to the same resolution as the reference data using nearest neighbour interpolation in order to match the pixel size of reference building footprints of 1m. All the layers under assessment were up-sampled in order not to favour a particular layer. The 13 selected cities for which the fine scale reference data was available are shown in figure 5.



Figure 5. Selected cities for the validation experiment.

The list of cities, reference data sources, year of download of reference data and the area of the specific site are shown in table 2. The second column refers to the extent of the validated area and not to specific city boundary area. This extent was determined according to the reference footprints coverage. The total validated area is 12586.34 km².

Table 2. Study sites with corresponding spatial extent, source of reference dataset and the year of reference dataset

City	Area [km ²]	Source	Year
Aizuwakamatsu	460.75	Open Street Map	2016
Glenorchy	96.25	Glenorchy City Council GIS	2013
Amsterdam	1846.5	Dutch National SDI (PDOK)	2016
Milano	227.25	Portale Cartografico Nazionale	2003
Montpellier	6756.84	IGN - BD Topo	2011
New York	857.75	NYC Open Data	2013
Novara	39.5	Portale Cartografico Nazionale	2003
Oslo	412.75	Open Street Map	2011
Surrey	329	Surrey City Council	2013
Torino	147.5	Portale Cartografico Nazionale	2012
Warsaw	928.75	Open Street Map	2011
Washington	154	Office of the Chief Technology Officer (OCTO)	2013
Dar es Salaam	329.5	Open Street Map	2016

4. Validation methodology

We introduce a validation framework that incorporates conventional methods of pixel-by-pixel accuracy assessment and analysis of grid-based differences in built-up in relation to settlements densities. The purpose of the framework is to achieve a comprehensive and systematic description of the accuracy and validity of the layers under comparison.

We first determine absolute accuracies and performances on the basis of a pixel-by-pixel accuracy matrix. We then explore the dependencies between the built-up and the physical settlement structures using a grid-based analysis. This is performed following two types of approaches: 1) analysis of the relationship between built-up densities derived from the reference data and the cumulative built-up area from the different layers; 2) correlation analysis between the sums of built-up pixels per cell derived from the different layers and the observed reference built-up cell sums. The rationale and the details of the three validation approaches are detailed in the following sections.

4.1 Performance metrics

This section is dedicated to the standard pixel-based accuracy assessment frequently used in remote sensing community. It is based on the metrics derived from confusion matrix (error matrix) (Congalton, 1991). For the analysis of absolute classification accuracies, we follow the recommendations by Foody (Foody, 2008) and base the interpretation of results on a combination of meaningful accuracy metrics beyond the use of a single statistic. As standard descriptive measures, we report the overall accuracy (OA), Kappa statistics, Commission and Omission Errors (CE and OE) according to the equations (1), (2), (3) and (4).

$$OA = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$Kappa = \frac{p_o - p_e}{1 - p_e} \quad (2)$$

$$CE = \frac{FP}{TP + FP} \quad (3)$$

$$OE = \frac{FN}{TP + FN} \quad (4)$$

Where the standard labels used in the binary confusion matrix are: TP = True positive; TN = True negative; FP = False positive; FN = False negative and $p_o = OA$;

$$p_e = \frac{(TP + FN) \cdot (TP + FP) + (FP + TN) \cdot (FN + TN)}{TP + FP + FN + TN}.$$

The overall accuracy informs about the correct classifications for all the pixels in a specific site, sum of diagonal members of the matrix divided by the total sum of pixels (equation 1). However, this metric is known for being vulnerable to bias from skew due to imbalanced data as in the case of built-up (Jeni et al., 2013). In the case of this analysis, the main interest is in the relative comparison of the different products. Therefore, despite the limitations of this metric, we have reported the overall accuracies of the different built-up layers that we analyse in combination with additional performance metrics. Kappa statistics introduced by Cohen (Cohen, 1960) highlights the differences between the actual agreement in the error matrix (i.e., the correctly classified sample units presented by the major diagonal) and the chance agreement presented by the column and row totals. Landis and Koch (Landis and Koch, 1977) proposed a categorization of Kappa in which values of 0.00 to 0.20 are regarded as poor, 0.21 to 0.4 as fair, 0.41 to 0.6 as moderate, 0.61 to 0.8 as substantial, and 0.81 to 1.00 as almost perfect agreement to ease the comparison of multiple classification outputs. CE (equation 3) is defined as the fraction of values that were classified as built-up but do not belong to that class. OE (equation 4) is defined as the fraction of values that belong to built-up but were not classified as such.

Using this pixel by pixel analysis means that for a particular study area, entire pixel population (entire extent of building footprints) is used in the experiment, that is, there was no sampling.

4.2 Built-up density analysis

The built-up density analysis introduced in (Ferri et al., 2014; Florczyk et al., 2016) is a grid-based method with a pre-defined cell size that allows assessing the relationship between the built-up density and the cumulative area of built-up derived from the different layers. This allows establishing a stronger understanding of the mapping capabilities of each sensor in relation to built-up density and structural characteristics. In this validation experiment, the cell size is set to 500x500 m. This cell size is a good compromise between the areas of the cities under analysis, the necessity to capture commission and omission errors and computational constraints. Several study sites are not big enough to use 1 km or bigger cell size (see table 2). Very small cell sizes, less than 50m, are close to pixel by pixel validation experiment and are demanding in terms of computation. For each cell, the sum of the built-up for each layer is calculated. The maximum cell sum is 250,000 m² because the common pixel size used for the resampled layers is 1m. In order to derive the built-up density of reference cells, the total sum of built-up pixels in a cell is divided by its maximum area. The explanation is given with a formula:

$$dens_i = \frac{\sum_{k=1}^N bu_k}{w_i \times h_i} \quad (5)$$

where, $dens_i$ is the density for a cell i , bu_k is the k built-up pixel in a specific cell i , N is the maximum number of built-up pixels in one cell, w_i and h_i are the width and the height of the cell, respectively.

The reference densities are ordered from the densest areas decreasing to the least dense. The cumulative built-up area sums are calculated for layers under validation respecting the order of reference densities. Since the reference cumulative densities are also calculated and plotted against the reference built-up (reference curve), all the other layers are compared to the reference curve.

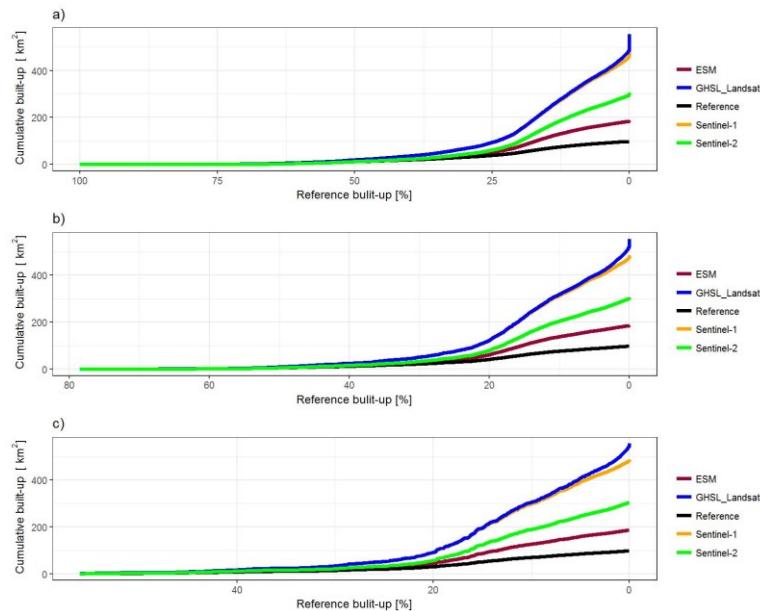


Figure 6. Comparison of different cell sizes for cumulative built-up curve analysis in Amsterdam study site: a) 250 m, b) 500 m and c) 1000 m. Reference (black line) is the “reference curve”.

Figure 6 displays the built-up density analysis for the city of Amsterdam with different cell sizes (250 m, 500 m and 1000 m). It shows that this analysis is robust to changes in the cell size.

The GHSL approach measures the percentage of built-up in a given spatial unit and this kind of analysis can explain whether the specific layer overestimates or underestimates the built-up area and at what particular densities.

4.3 Regression analysis of built-up

The regression analysis uses similar input as the cumulative built-up density analysis by exploring the correlation between the sums of built-up pixels per cell derived from the different layers and the observed reference built-up cell sums. We analyse the scatterplot matrices displaying the pairwise correlations. To support the analysis of scatterplots, we also compute the Pearson coefficient of correlation r , the slope and the intercept from first order linear regression:

$$y = a \cdot x + b \quad (6)$$

where: a is the slope coefficient and b is the intercept, y correspond to specific layer under analysis (ESM, S1, S2 or GHSL-Landsat) and x to the reference layer derived from building footprints.

Through this analysis, we explore the possibility of estimating the “actual” built-up area given an input layer derived from one of the satellite sensors assessed in this work. The outputs of the correlation analysis can also give indication on the degree of interoperability between the different built-up layers which may facilitate the task of sharing of data between different agencies and organizations working on human settlements.

5. Results and discussion

The proposed validation framework is applied to the 13 study cities. The results are structured as follows:

- All European sites (7 in total) are presented together with the ESM layer included in the assessments,
- All 13 sites are presented (European and non-European) excluding the ESM layer which is only available in Europe.

5.1 Results of performance metrics

Figures 7 and 8 display the comparative performance metrics per city (In figure 7 only European cities are shown with ESM layer included in the analysis, while in Figure 8, all the cities are presented but excluding the ESM layer).

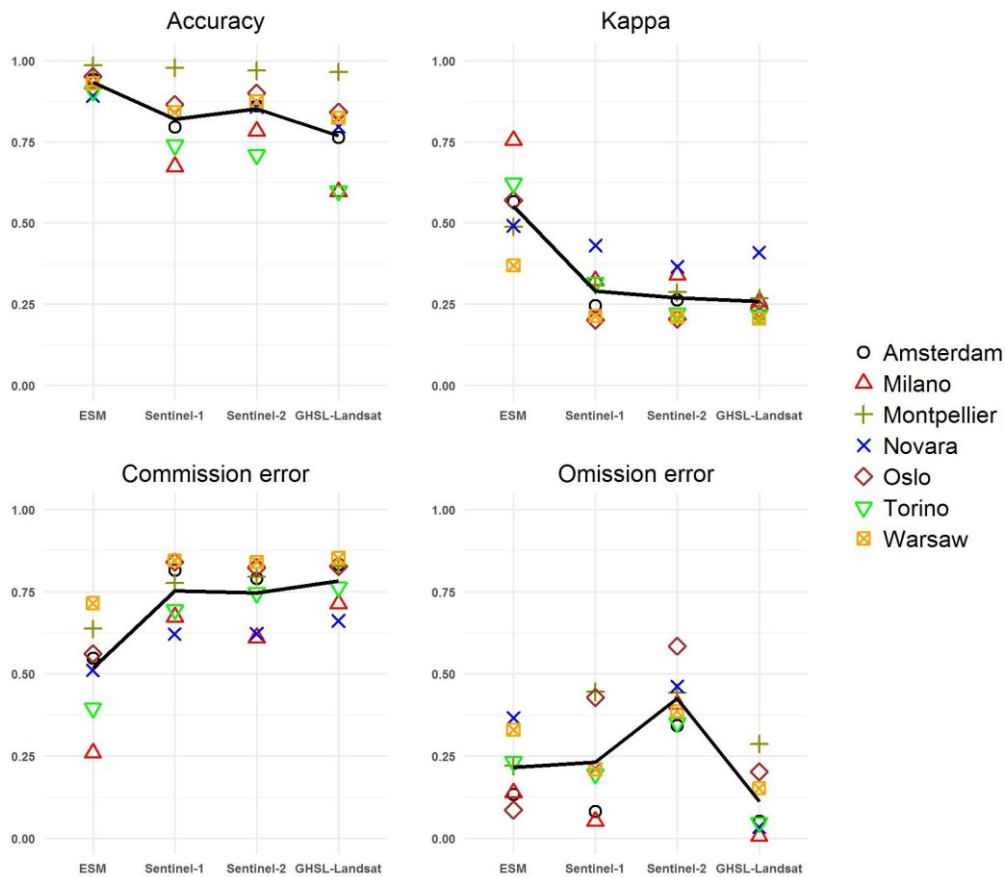


Figure 7. Performance metrics for European cities. Black lines correspond to the average value of a given measure.

The analysis of the four different performance metrics shows that the ESM resulted in the highest values of overall accuracy and kappa and the lowest value of CE. The performances of S1 and S2 in European cities are very similar. S2 produced a slightly higher value of average OA but the average kappa value of S1 is marginally higher. The CE of GHSL-Landsat is the highest, but in general, all four layers show high average commission errors greater than 0.5. In terms of OE the GHSL-Landsat shows the best results with very low values, even lower than ESM. However, low OE is not necessarily synonym to good classification results (Wenkai Li and Qinghua Guo, 2014). If the classifier is only predicting positive or negative instances, its kappa value will be 0 with no omission or commission error. The OE of S2 is significantly higher than OE of the other layers. This may be explained by the fact that the S2 workflow includes a textural analysis aimed at refining the derived built-up layers by removing roads and other non-built up areas with high confidence values. The textural refinement seems to exclude a lot more than non-built-up areas by also excluding large buildings (e.g. industrial) and some groups of buildings in very dense cities (figures 16, 17 and 22).

Performance metrics for all cities (figure 8) show that S2 slightly prevails with the highest average OA and Kappa values and the lowest OE. GHSL-Landsat has the highest CE. This is because GHSL-Landsat tends to classify some road networks, bare soil in arable land and river beds, etc. as built-up (figures 16, 17 and 18). S1 shows overall also good performances in between those of S2 and GHSL-Landsat.

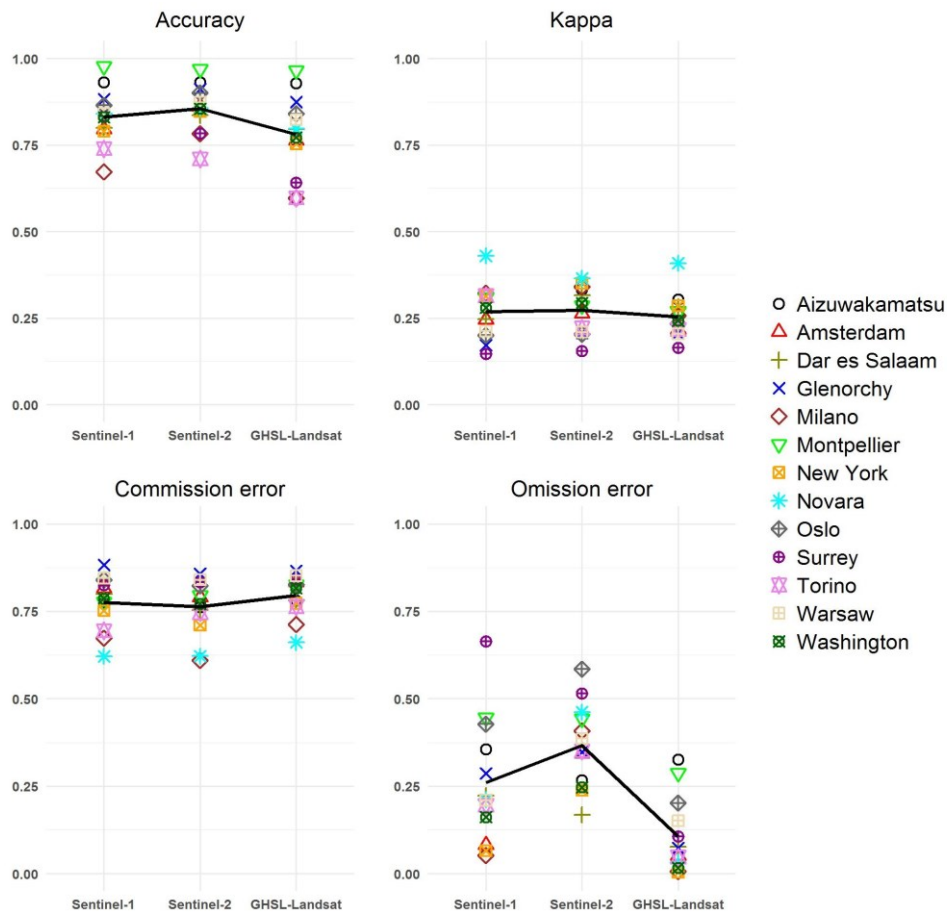


Figure 8. Performance metrics for all cities. Averaged values by city are represented with black line.

5.2 Results of performance metrics

In figures 9 and 10, the results of the built-up density analysis are presented showing the behaviour of the cumulative built-up area of different layers, including the reference building footprints (in black) in relation to the reference built-up densities. The densities calculated per grid cells of 500x500m are ordered from the densest to the least dense. The cumulative built-up sums are calculated for the layers under validation and plotted respecting the order of reference densities. The analysis is made by visual comparison of the curves derived from ESM, S1, S2 and GHSL-Landsat to the "reference curve".

Figure 9 displays the cumulative built-up curves for European cities. Since these are cumulative areas of built-up, the maximum value is the total area of built-up detected by the layers. Evidently, the curves calculated with the ESM are the closest to the reference curves both in terms of shape and distance. It is notable that for Milano and Torino the cumulative built-up curves for ESM almost coincide with reference layer. Hence, the final estimated area is close to the reference data. However the ESM still tends to overestimate in general the area of built-up, except in the case of Torino where it slightly underestimates the built-up area. The almost perfect match between the ESM and the reference curve especially in the case of Torino and Milano can be explained by the fact that the ESM workflow uses the same reference building footprints as an auxiliary dataset for filling the gaps in the final ESM product. This certainly introduces a bias in the validation.

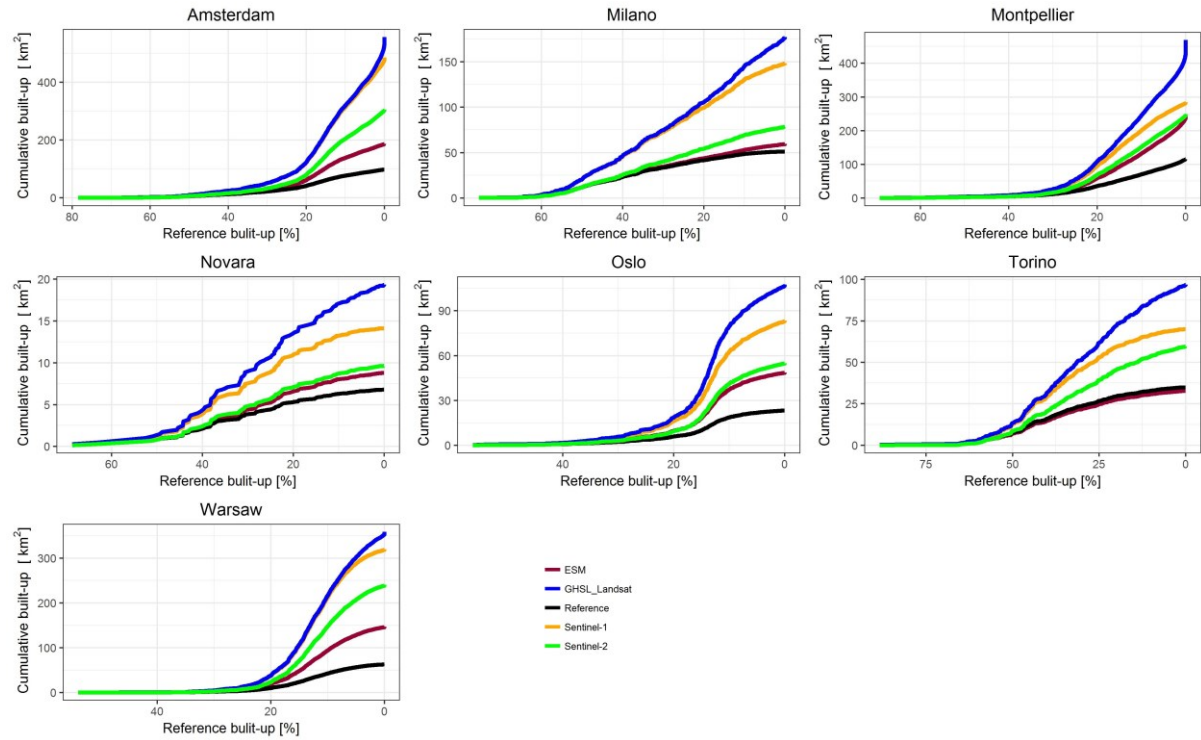


Figure 9. Cumulative built-up curve analysis for European cities. The x axis corresponds to decreasing reference built-up densities calculated for grid cells of 500x500m. Y axis is the cumulative built-up for all layers, including the reference layer.

The GHSL-Landsat is overestimating the area of built-up significantly and this overestimation is reached very fast. S2 shows good results with a curve very close and parallel to the ESM curve (e.g. in the case of Montpellier, Oslo, Novara and Milano). S1 performed better than GHSL-Landsat but also has a general tendency to overestimate the built-up area in sparse urban zones.

For the cities of Warsaw, Oslo, Amsterdam and Montpellier the curves start to diverge when the reference built-up density is around 20%. At that point the cumulative areas for all layers start to differ greatly and that's when we begin observing the overestimation. The reference data in those cities covers not only the dense city centres but also sparse built-up and some rural zones. In those cases, it is possible to detect the tendency of the different layers to over-detect the area of built-up in sparse and scattered settlement patterns. In the Italian cities (Novara, Torino and Milano) the reference data covers essentially the urban core areas where the density of the built-up is very high. In that situation, all the built-layers show almost the same behaviour and start overestimating around 50% of built-up density.

Figure 10 presents the cumulative built-up curves for all the cities in the validation study. Similarly to the situation in Europe, the results show an overestimation of built-up areas for all three layers (S1, S2 and GHSL-Landsat). This is normal due to the nature of the built-up observed from the satellite sensors and to the layers inherent semantic definition of settlement areas that does not comply with individual building outlines.

There are two special cases in the results. The first one is in Aizuwakamatsu (Japan) where the S2 produced a total area of built-up higher than in the one estimated from Landsat (see figure 11). This site is characterized by many rural areas which surround the city. Unlike for the other sites, S2 OE was 0.27, which is lesser compared to Landsat (0.33) and S1 OE (0.36) (figure 8), due to small number of large buildings. Besides, smaller scattered settlements detected by S2 were rarely or not at all detected by Landsat sensors. Visual examples are provided in the figures 19 and 20.

The second special case is in Surrey (Canada). The city is composed mostly out of residential houses with characteristic dark roofs made of asphalt shingles that make them difficult to distinguish from roads with optical data (see figure 24). In the case of this city, the S2 was outperformed by the S1 which captured those particular building types resulting in a very good match with the reference cumulative built-up curve.

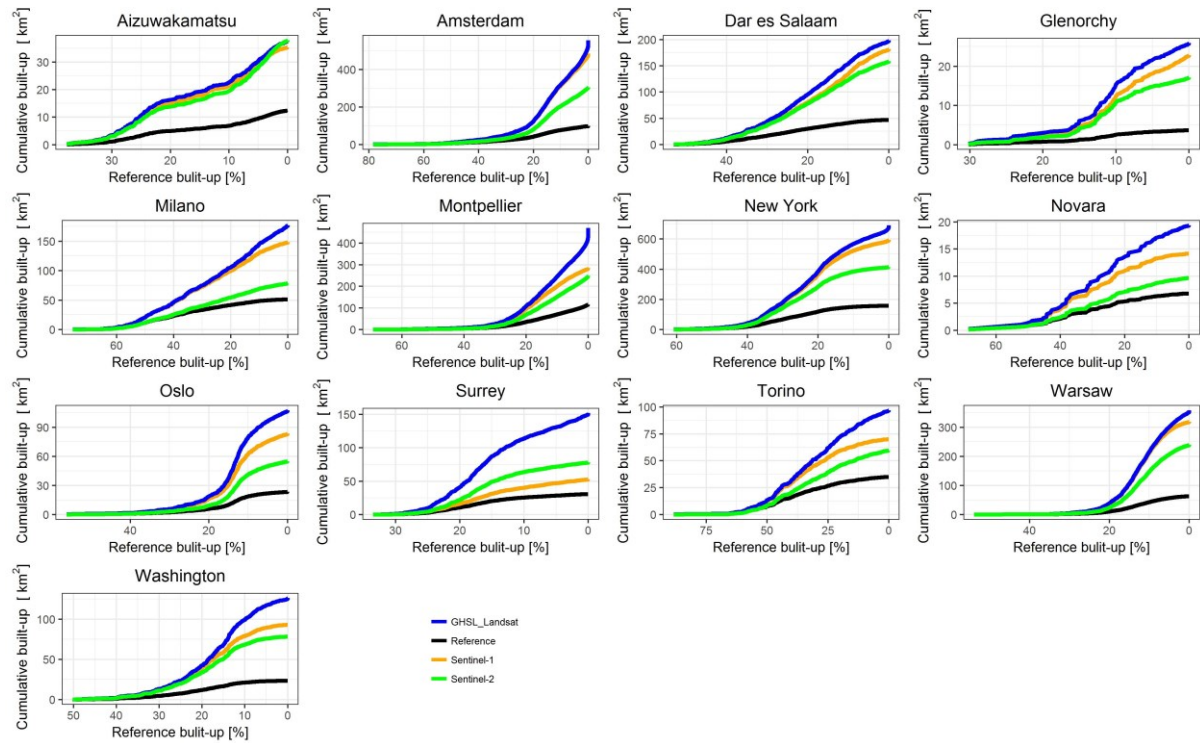


Figure 10. Results of the cumulative built-up curve analysis for all the study sites.

Figure 11 provides the summary of the estimated built-up area for each city and for each layer as well as the reference built-up area. The results highlight a similarity in the total area of built-up as derived from SPOT5/SPOT6 (ESM) and Sentinel-2, especially for the cities of Montpellier, Novara and Oslo. S1 and Landsat also provided high similarity of the total estimated built-up for several cities: Glenorchy, Novara, Dar es Salaam and Warsaw.

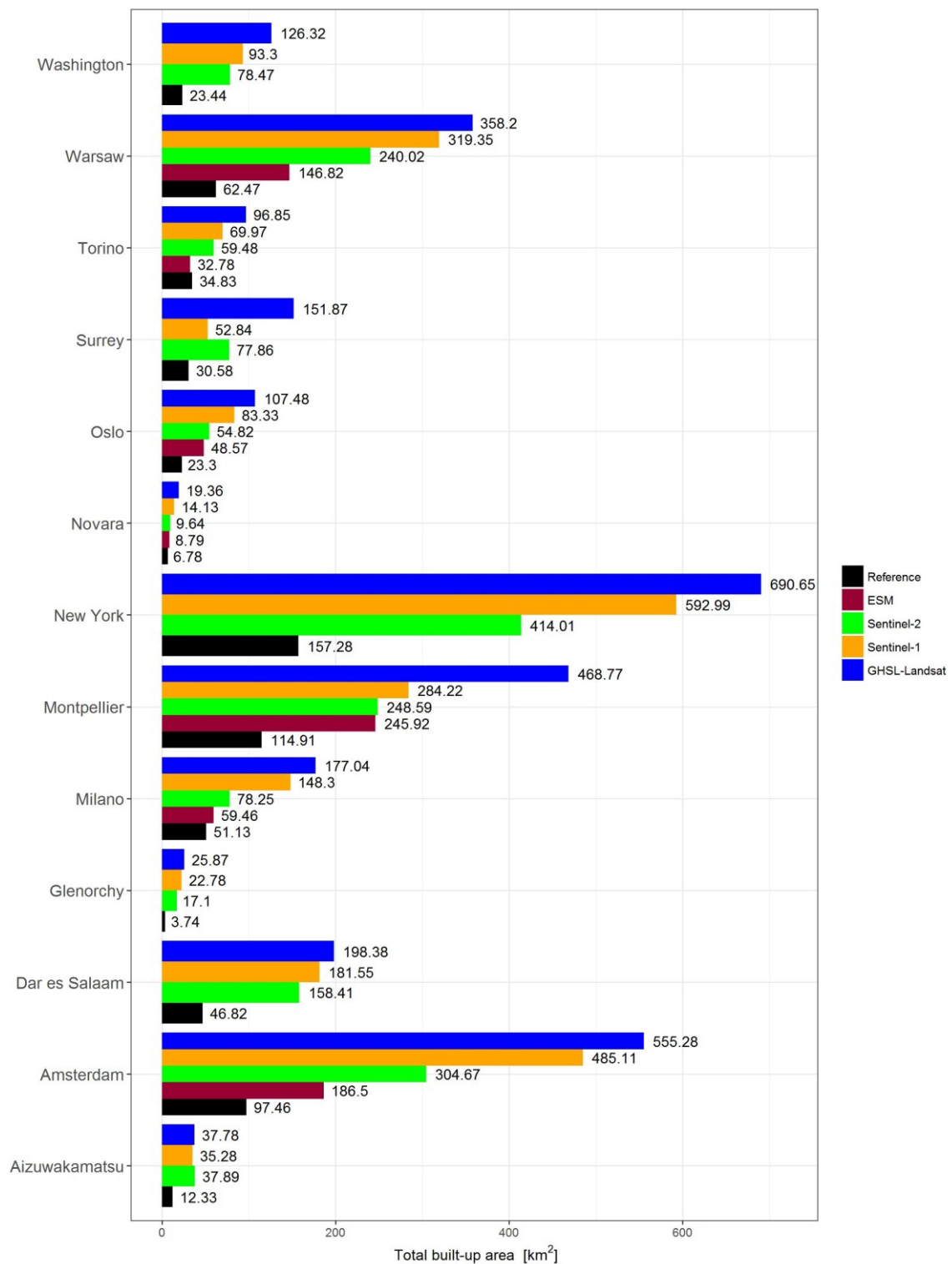


Figure 11. Estimated built-up area per layer and reference built-up area for all sites

5.3 Results of regression analysis of built-up

The values of the scatterplot matrix on the x axis correspond to the reference built-up, whilst the values on the y axis are the values of built-up for the layers under validation (ESM, S1, S2 and GHSL-Landsat). Figure 12 displays the correlation matrices (scatterplot matrices) for the European cities.

The results confirm the cumulative built-up curve analysis implying a more accurate representation of built-up structural variability by the ESM in comparison to the other

layers. This is manifested by the high r values ranging between 0.91 and 0.96. Also the linear relationship is very prominent between the ESM and reference footprints especially in the case of Milano and Torino.

From the scatterplot, it can be also retained that both the GHSL-Landsat and S1 over-detect built-up with the GHSL-Landsat showing somehow a second order polynomial relation with the reference built-up area: the values in the GHSL-Landsat vs Reference scatterplot (first column, fifth row) are mostly concentrated in the left and upper left corner of the plot. That is, for low values of reference built-up densities; the GHSL-Landsat gives much higher values than the other layers. Significant numbers of values are sealed at the top. This relationship is similar also in the case of GHSL-Landsat-ESM, S1, S2 plots (fifth row, columns two, three and four), and values are concentrated in the upper parts of the scatterplot. This is the consequence of many of the GHSL-Landsat cells which have 100% built-up density as opposed to the ground truth data.

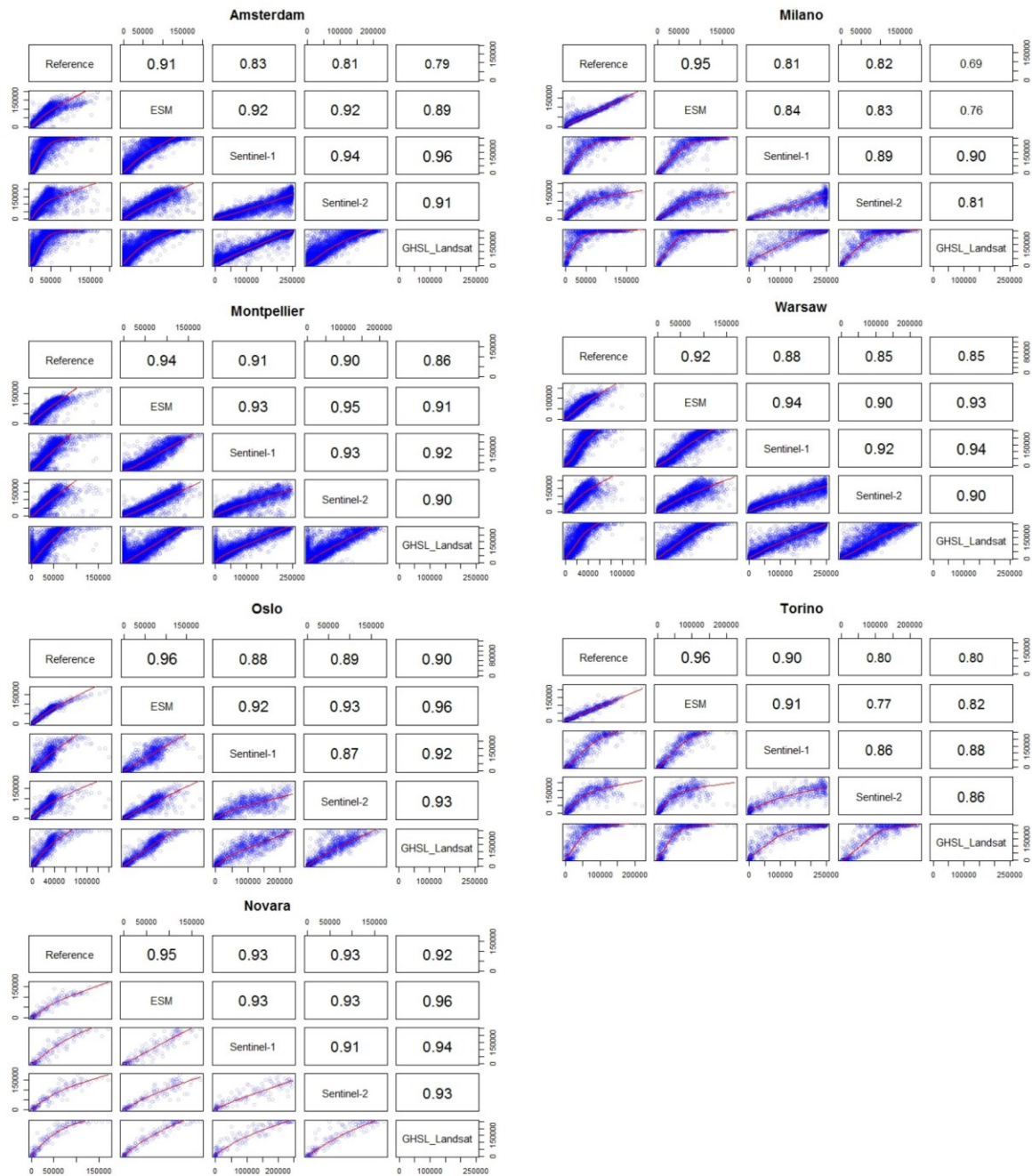


Figure 12. Scatterplot matrices with correlation coefficients for European study. X axis corresponds to reference built-up extracted from 500x500m grid cells, whilst Y axis corresponds to built-up area obtained from grid cells for specific layers (ESM, S1, S2 and GHSL-Landsat)

Overall, when looking at the results obtained for all study sites, we notice that the r values are the lowest in the case of GHSL-Landsat for 9 out of the 13 validated sites. However, for the sites characterized by significant presence of non-built-up and sparse built-up cells (e.g. Oslo, Glenorchy, Surrey and Aizuwakamatsu), the GHSL-Landsat performs very well in terms of r values. That is, Landsat performed well in not detecting the non-built-up. For cities with very high building densities the GHSL-Landsat gave the smallest coefficients of correlation, only 0.72 and 0.69 (New York and Milano).

Both the S1 and S2 are strongly correlated with reference layer and with the ESM. Similarly to the GHSL-Landsat, they tend to overestimate the built-up area, especially S1. It is important to note that the shape of the relation between S2 and the reference is

different than that observed for S1 and GHSL-Landsat: there is less overdetection in the S2 especially in fragmented arrangements as in the case of Montpellier and Amsterdam. These results confirm those obtained with the cumulative built-up curve analysis.

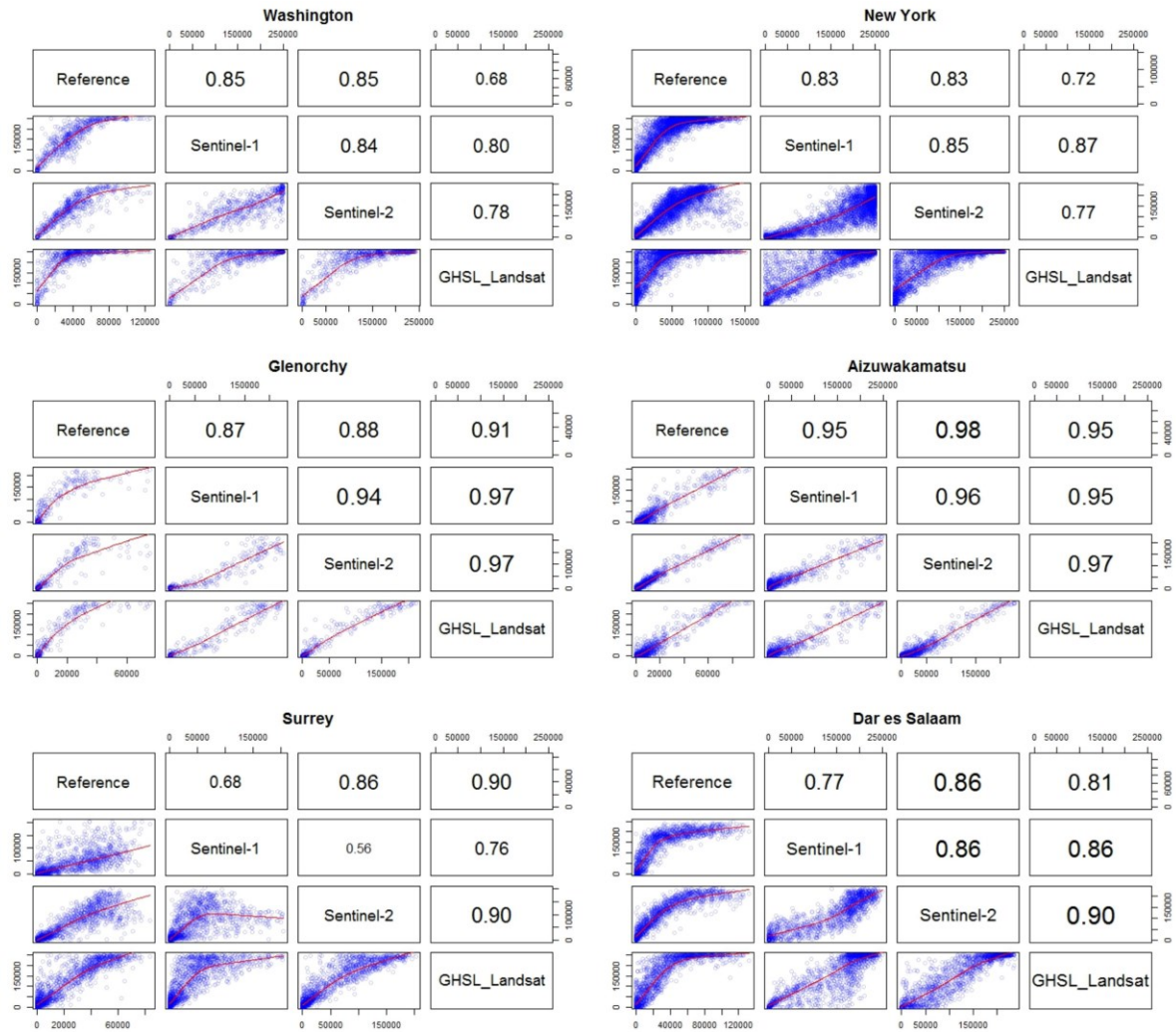


Figure 13. Scatterplot matrices with correlation coefficient in non-European cities.

The slopes of the regression analysis are displayed in figures 14 and 15. For the ESM, the slopes are stable and fall in the range [1, 2]. Given the estimated area of built-up obtained with the ESM (Figure 11) in comparison to the actual reference built-up area, and considering the average slope value of the fitted linear models, it is reasonable to state that the ESM overestimates the built-up area by a factor of 2.

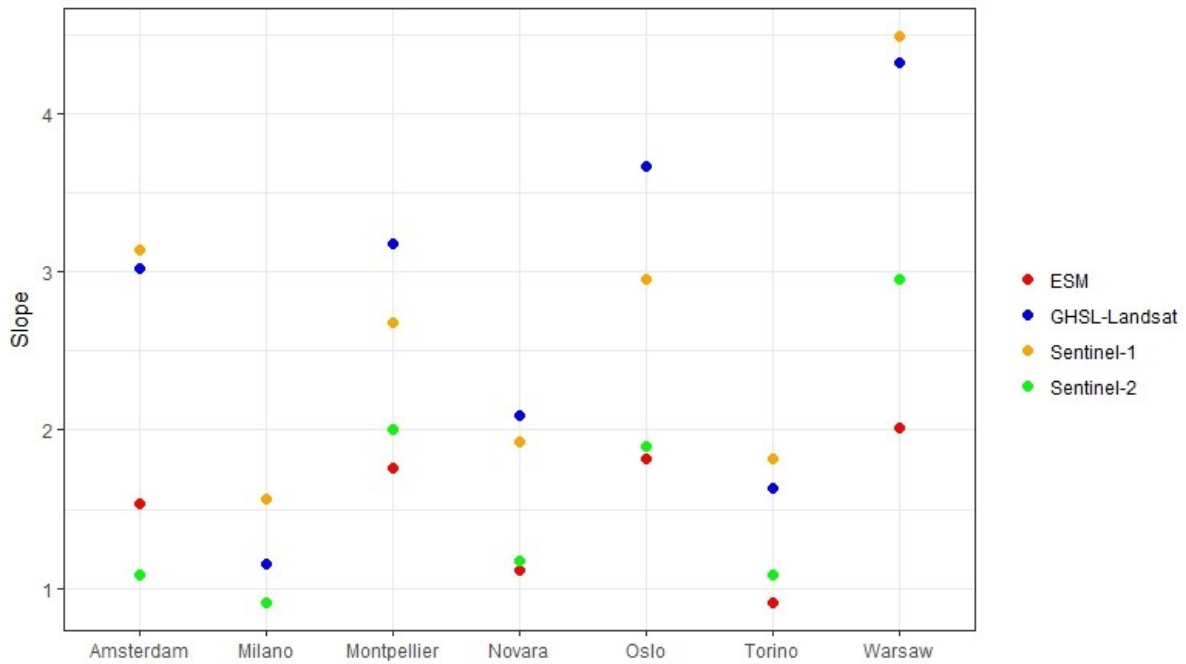


Figure 14. Slopes from the linear regression model plotted for each European city and for each product.

It is interesting to note that slopes for S2 for 4 European cities are also in the range [1, 2] except for Warsaw. The average slope value for S2 in the case of all cities is 2.08. Considering also a linear regression model, it is possible to suggest that the S2 overestimates the built-up area by a factor of 2.08. In the case of S1 and GHSL-Landsat, the slopes show a much larger variability making it difficult to derive meaningful average values of slopes. However, we notice that in rare cases (e.g. Surrey, Dar es Salaam, New York) the slopes obtained with S1 are lower or very close to those obtained with S2. The large variability in the slope values, especially in the case of S1 and GHSL-Landsat suggest that it is difficult to build a model that covers all the different cities. It would be more appropriate to consider separately the different landscapes (e.g. by continent) and the different settlement patterns (e.g. dense cities, fragmented cities, rural areas) and attempt to propose a model per each context. However, to be able to propose such a model, a large sample size is needed covering a large diversity of cities stratified by landscape and settlement patterns.

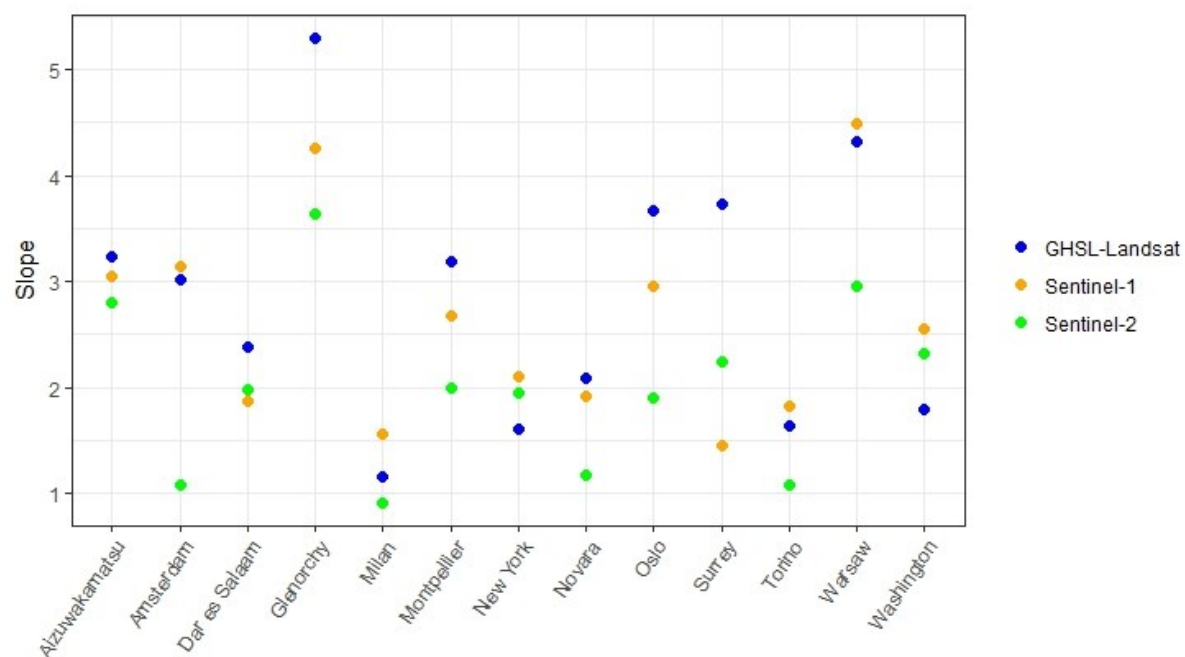


Figure 15. Slope values plotted for each validated site.

In the Annex, a summary of the coefficients (slope and intercept) of the linear regression models are provided.

6. Visual inspection

To support the results of the quantitative validation presented in the previous chapter, some examples of visual inspection of the classification results are given here. They allow understanding in which situations and contexts the omission and commission errors may occur and where improvements in the information extraction workflows are needed.

Figure 16 provides a comparison of SML classification results for GHSL-Landsat, S1 and S2, overlaid on satellite imagery obtained from Google Earth. It can be seen that GHSL-Landsat classifies roads (runway) and parking lots as built-up. In the case of S1, roads are not classified as built-up while parking lots are falsely detected in the built-up class. S2 correctly omits roads and parking lots from the detected built-up but fails in classifying large buildings.

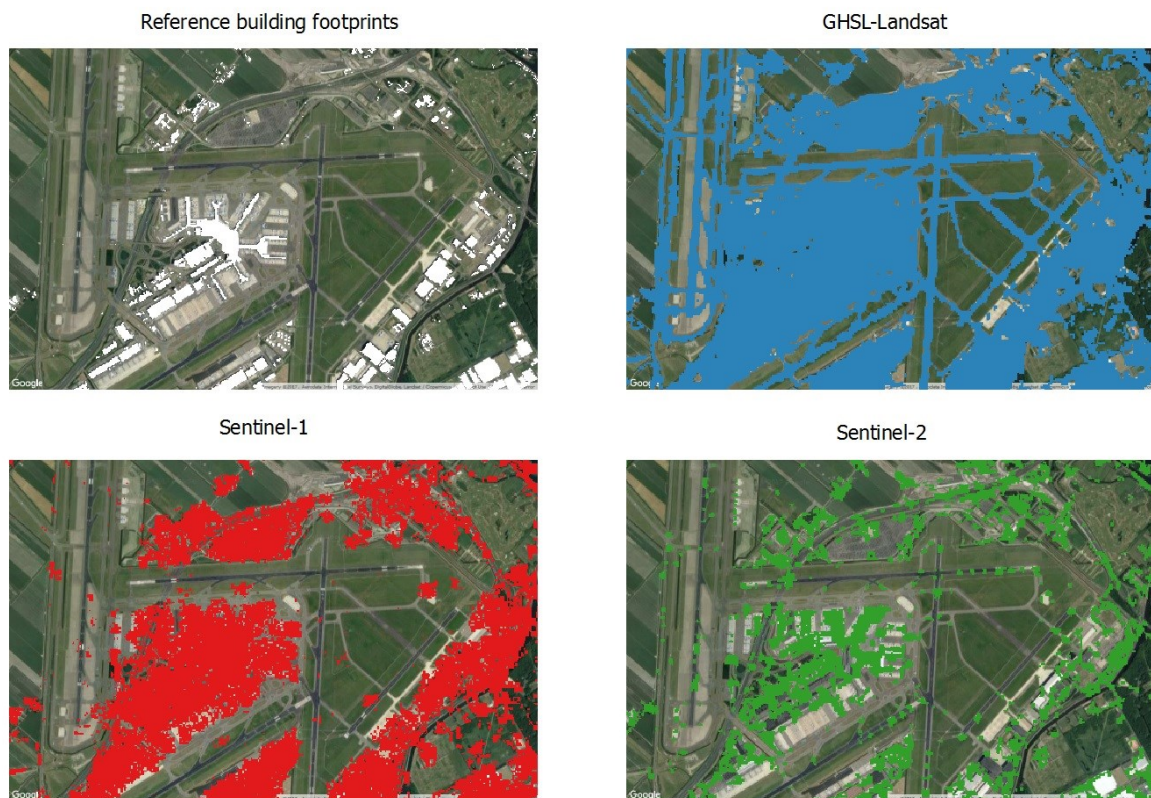


Figure 16. Close view of built-up derived from S1, S2 and Landsat in Amsterdam (location: 52.30N, 4.76E).

Another example shown in figure 17 confirms the omissions of S2 in case of large buildings and the commission errors for GHSL-Landsat in the case of parking lots, grasslands and bare land.

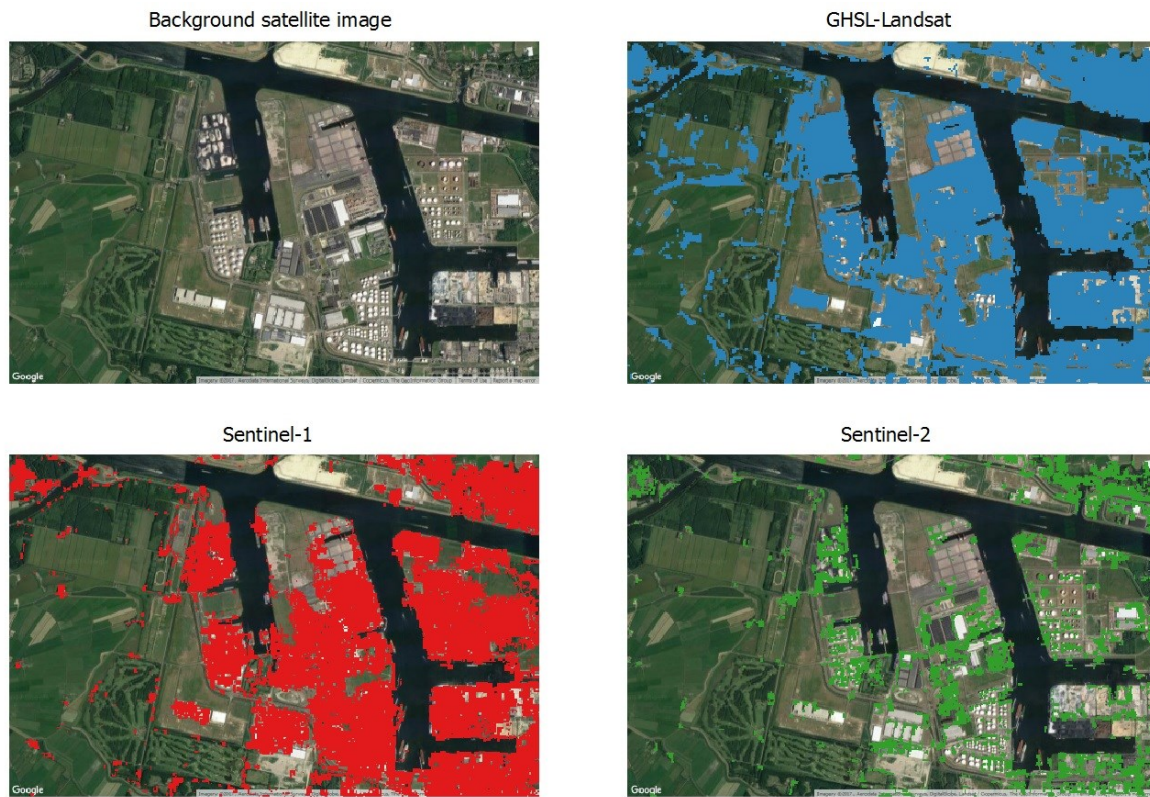


Figure 17. Close view of built-up derived from S1, S2 and Landsat in Amsterdam (location: 52.41N, 4.77E).

The following example (figure 18) proves that the GHSL-Landsat also tends to classify many agricultural fields as built-up. This is partially improved with S1 radar data. The best outputs in terms of reduction of commission errors in agricultural fields are those obtained with S2. It is also evident that S2 outputs allow a better characterization of the porosity (i.e. amount of open spaces) within the built-up area compared to S1 and GHSL-Landsat. Balancing omission errors and the capacity to identify open spaces is an issue requiring careful attention and additional experiments in the prototyping phase of the S2 workflow.

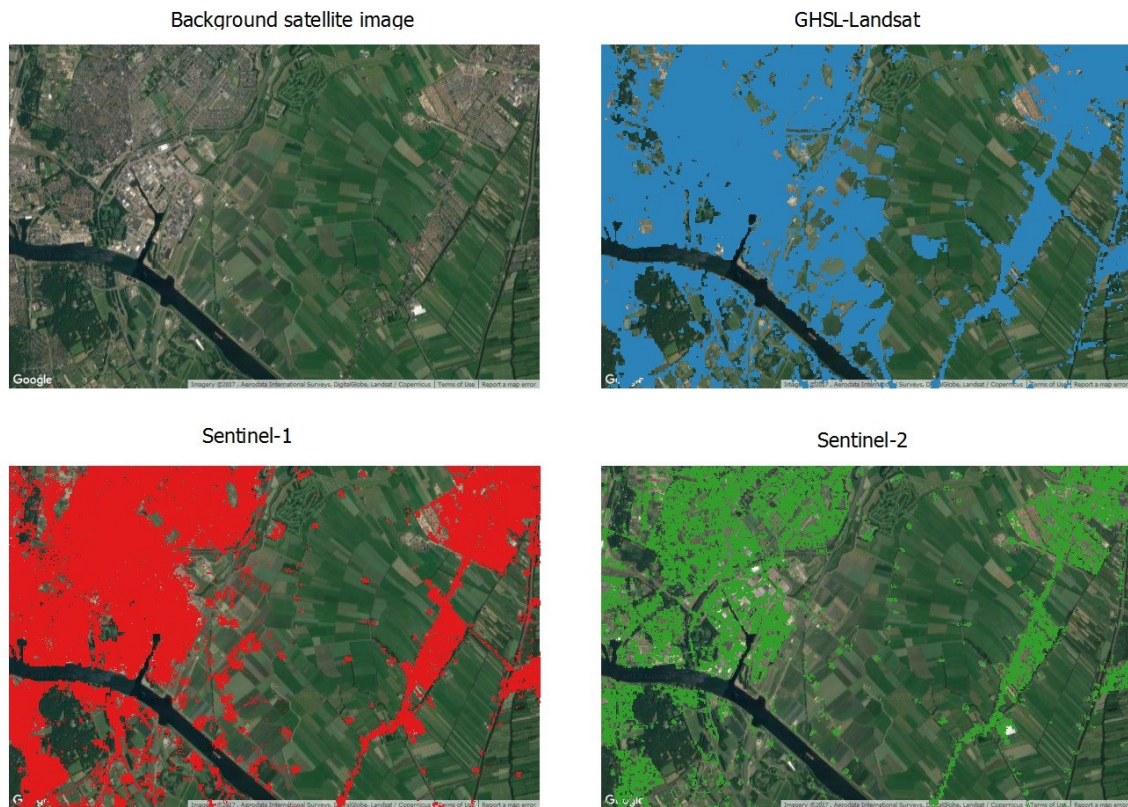


Figure 18. Close view of built-up derived from S1, S2 and Landsat in Amsterdam (location: 52.45N, 4.70E).

Examples for Aizuwakamatsu case study are shown in figures 19 and 20. The focus here is on the capacity of the different sensors to detect sparse built-up. It is interesting to note that S2 is capable of detecting very small scattered settlements with few commission errors when GHSL-Landsat and S1 are failing.

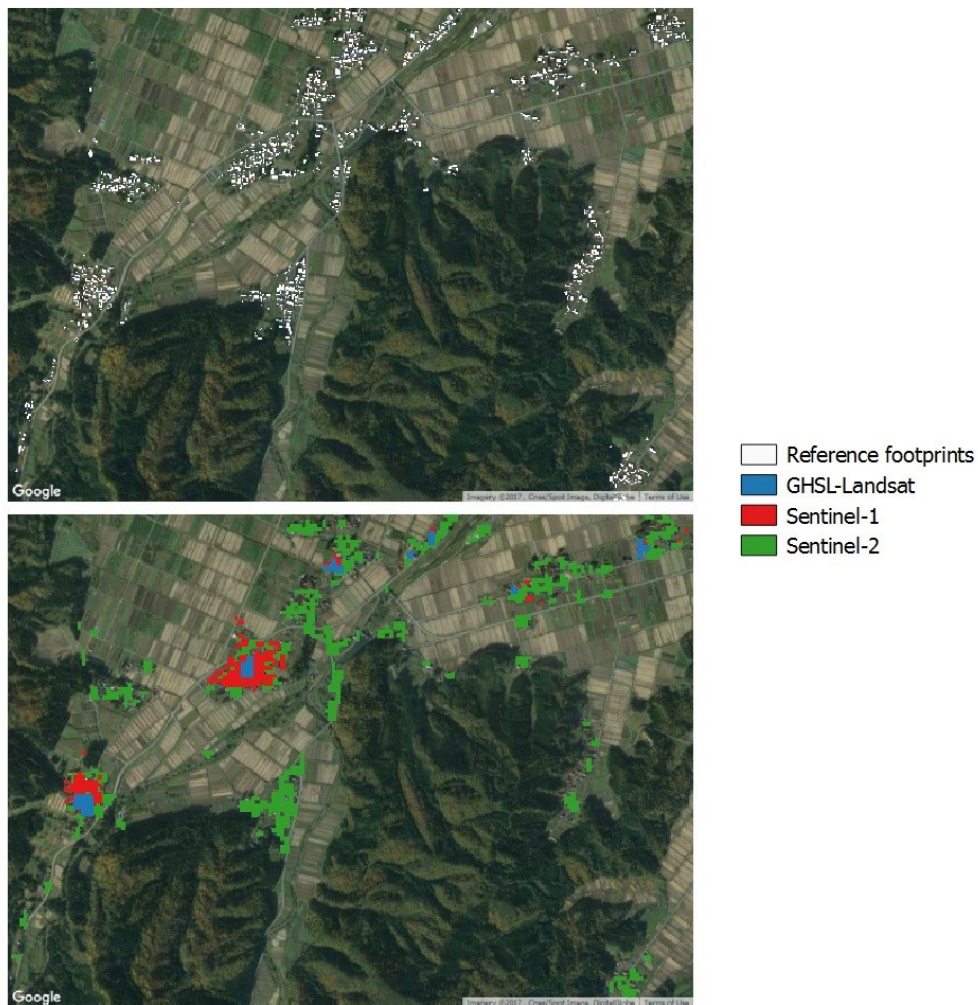


Figure 19. Close view of built-up derived from S1, S2 and Landsat with reference footprints in Aizuwakamatsu (location: 37.42N, 139.81E).

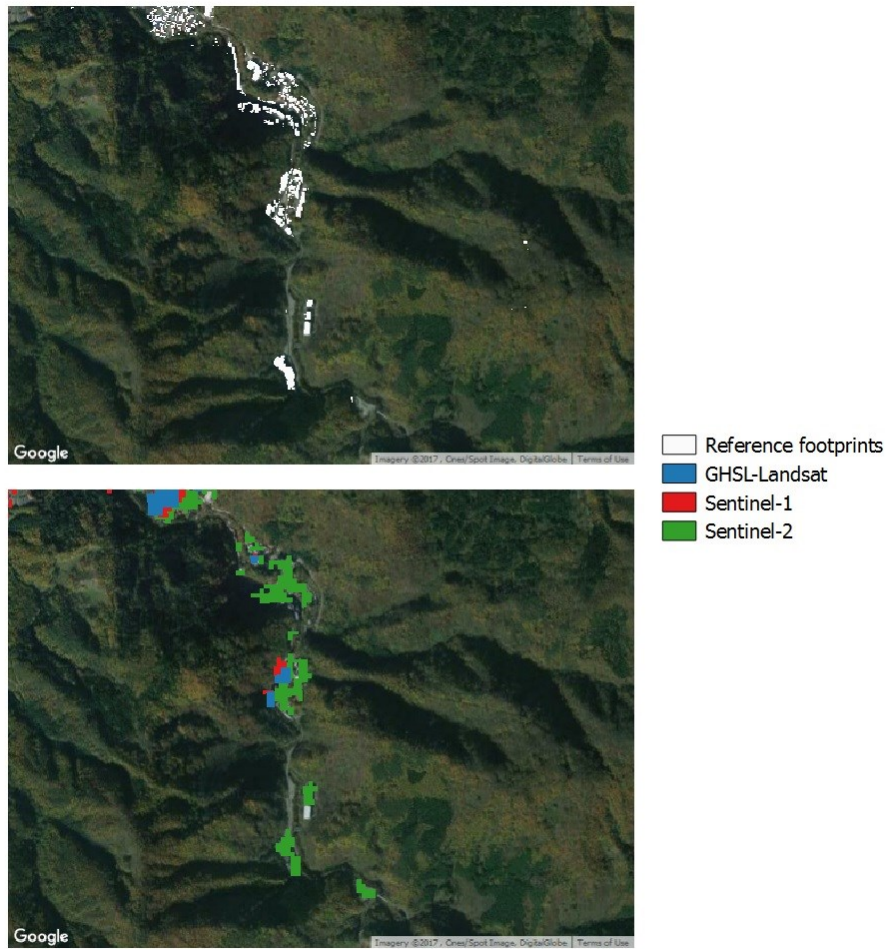


Figure 20. Close view of built-up derived from S1, S2 and Landsat with reference footprints in Aizuwakamatsu (location: 37.47N, 139.96E).

Sandy beaches can introduce high commission errors along the coast in optical data because of the sand high reflectance properties which are similar to man-made objects. An example is shown in figure 20 where the S1 and S2 succeed in excluding sand beaches from the built-up class as opposed to GHSL-Landsat. Also the S1 and S2 outputs are more refined with a better characterization of open spaces in comparison to GHSL-Landsat which provides more compact and clustered built-up blobs.

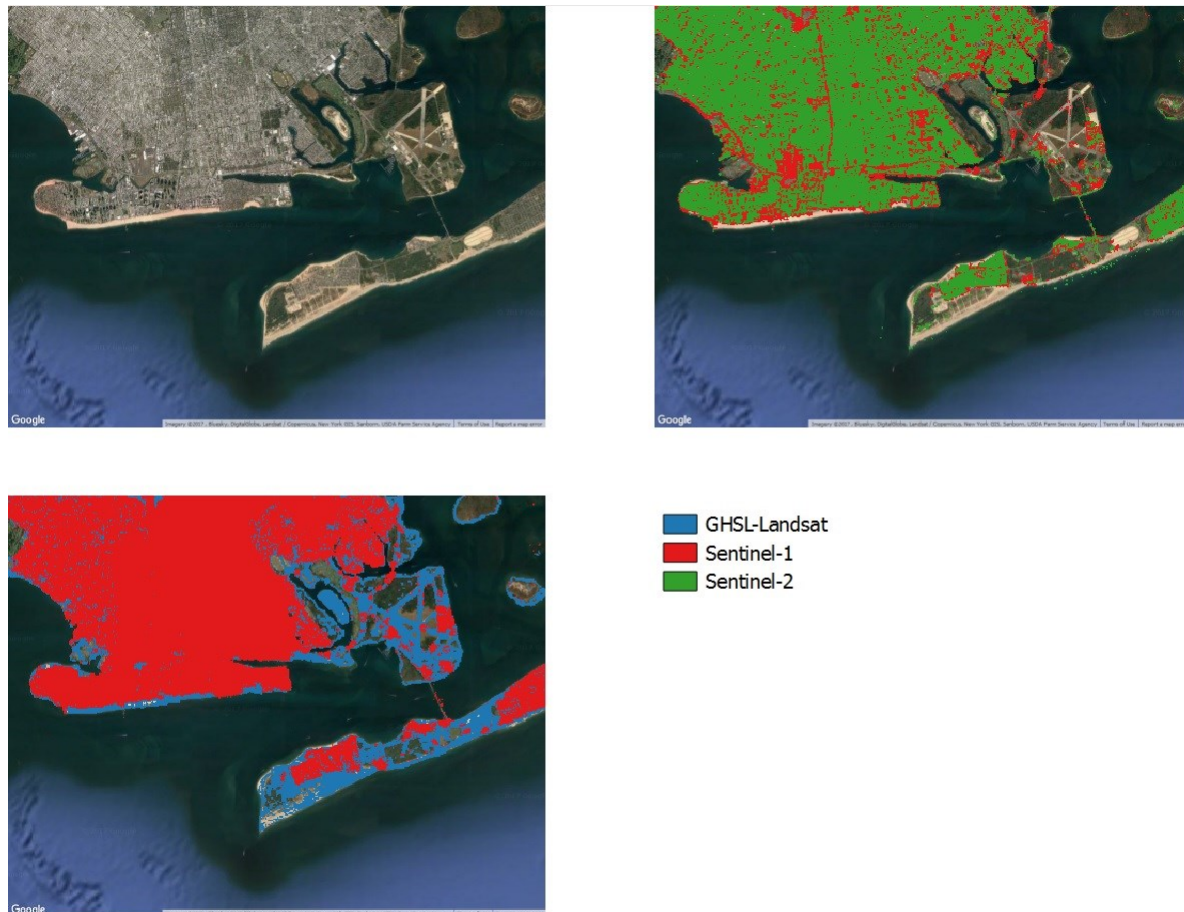


Figure 21. Close view of built-up derived from S1, S2 and Landsat in New York (location: 40.57N, -73.95E).

Although S2 gave higher overall accuracy than S1 and Landsat and lower commission errors, its built-up omissions are significantly higher. The following figure illustrates the excessive S2 omissions of large industrial buildings for Amsterdam study site.



Figure 22. Close view of built-up derived from S2 in Amsterdam (location: 52.23N, 4.74E)

Figure 23 shows an example of the level of precision of the built-up as described in the ESM layer. The similarity/consistency between the ESM and building footprints is very striking. Also, some of the buildings missing from the reference data are detected by ESM layer.



Figure 23. Close view of built-up derived from the ESM in Oslo (location: 59.92N, 10.74E).

Figure 24 represents high resolution satellite imagery overlaid with GHSL-Landsat, S1 and S2 in Canada study site (Surrey). Site is characterized with many residential houses with significant presence of vegetation and shadow. Furthermore, the residential houses have a specific orientation which can introduce difficulties for S1 built-up detection.

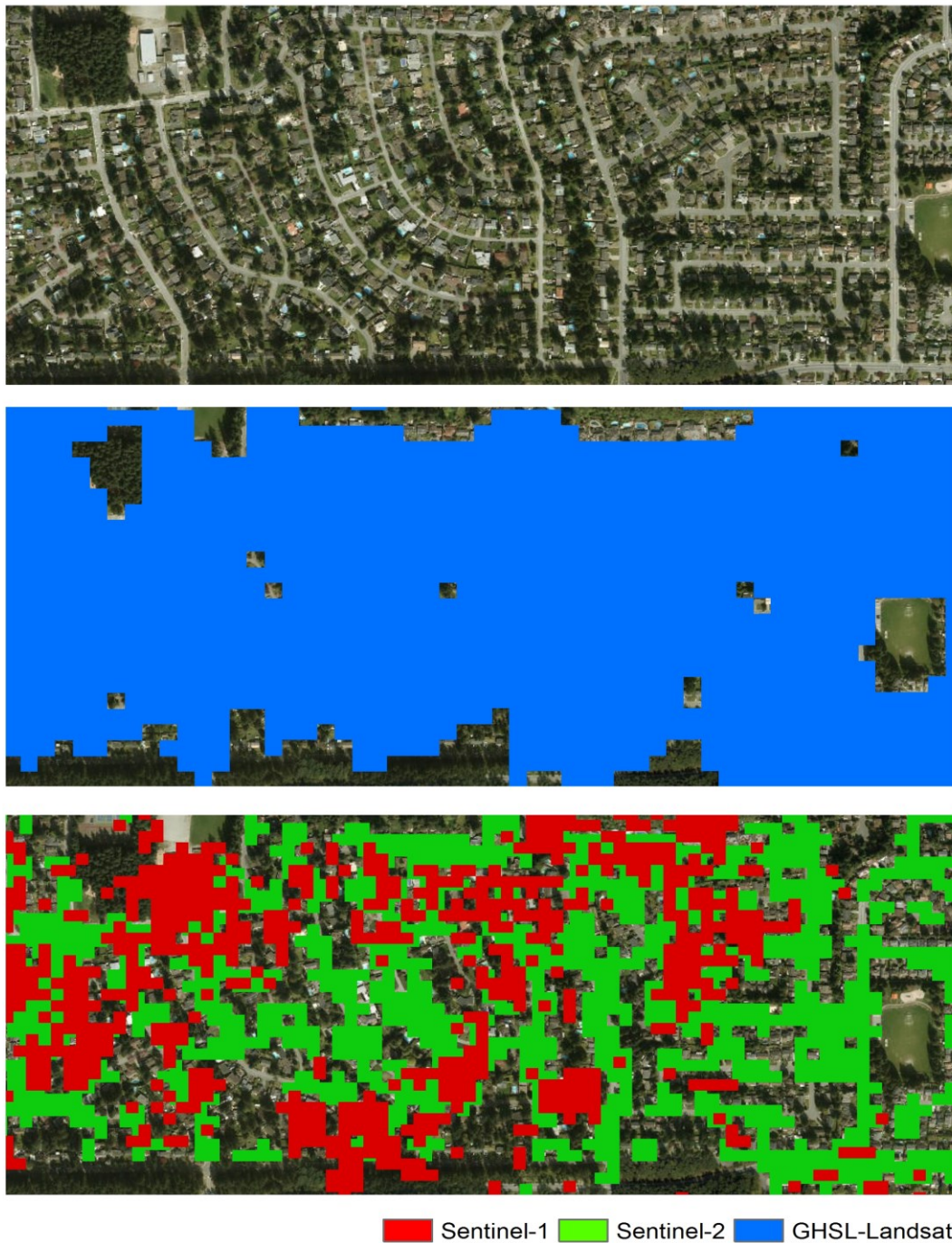


Figure 24. Close view of built-up derived from S1, S2 and Landsat in Surrey study site (location: 49.12N, -122.91E).

7. Conclusion

An inter-sensor comparison of built-up areas derived from different sensors was presented for 13 selected cities and 4 built-up products obtained in the framework of GHSL. The quantitative validation was performed using detailed building footprints available from OSM or from national mapping agencies and local authorities. We first determined absolute accuracies and performances based on a pixel-by-pixel accuracy matrix, considering the entire population of reference pixels. We then explored the dependencies between the built-up and the physical settlement structures using a grid-based analysis. This was accomplished by using two types of approaches: 1) analysis of the relationship between built-up densities derived from the reference data and the cumulative built-up area from the different layers; 2) correlation analysis between the sums of built-up pixels per cell derived from the different layers and the observed reference built-up cell sums. Finally, a visual comparison of built-up classification results was performed illustrating areas of agreement/disagreement between the layers.

This analysis allows deriving the following main observations:

- i) ESM shows the most accurate results relative to the reference data;
- ii) Both S1 and S2 showed an improved extraction of built-up compared to Landsat;
- iii) All the layers overestimate the built-up area because of their inherent semantic definition of settlement areas that does not comply with individual building outlines. In particular, the GHSL-Landsat with the lowest spatial resolution, showed to be the product with the highest overestimation of built-up areas.
- iv) The results varied significantly across the different type of cities suggesting the need to group the analysis per types of landscape and settlement patterns;

Based on all the validation experiments the ESM showed the most accurate representation of the built-up independently from the settlements patterns and their densities. Given that reference building footprints are rarely available for large scale validation, the outputs of this study suggest that ESM represents a good alternative for the validation at the European level. Besides, previous validation works of the ESM (Florczyk et al., 2016) showed that this layer generated by automatic image information extraction achieves 96% of agreement with the LUCAS dataset. The omission and commission errors are less than 4% and 1%, respectively. This makes it a good reference for the validation of the GHSL-Landsat, S1 and S2 within Europe.

Compared to GHSL-Landsat, S2 and S1 showed significant improvements associated with the exclusion of agricultural fields, parking lots, sand beaches and roads from the detected built-up. S2 was good in detecting small scattered settlements but failed in detecting large buildings in dense urban zones. Conversely, while S1 did not succeed in identifying scattered settlements, but correctly classified large industrial buildings. These results outline complementarity between S1 and S2 sensors for increased accuracy in the detection of built-up. Deeper insights into the characterization of human settlements can be certainly gained from the integration of S1 and S2 results.

The large variability of the results of the validation between the study sites, suggest the need for expanding the sample size to cover cities in different landscapes and with different settlement patterns. The results, which would incorporate a wide scattered sample size, could be used for developing robust cross sensor built-up models. Future works will be focused on a systematic and exhaustive consistency check of the different products and extensive validation in view of the development of sensor-specific models for deriving the “actual” areas of built-up from the different available built-up products

This report showed that different validation experiments combined with visual examples can contribute to a better understanding of the similarities and the disparities between the different built-up information layers. It also shed lights on the potentials for exploiting the synergies between the sensors for consistent mapping of human settlements at the regional and global scales.

Acknowledgements

The authors wish to thank the GHSL team who contributed in several ways to this work mainly:

Aneta Florczyk, Panagiotis Politis, Luca Maffenini, Daniele Ehrlich, Matina Halkia, Donato Airaghi, Martino Pesaresi and Thomas Kemper for their advices and the technical and scientific support. We would also like to extend our gratitude to Dr. Dušan Jovanović, from the Faculty of technical sciences in Serbia, for reviewing this report.

Annex

Tables 4 and 5 provide the summary of coefficients (slope and intercept) derived from linear regression as well as their standard errors for all sites and for all layers. All the coefficients are significant at 0.01. Even though slope values are not very high in Milano and Torino case for S1, they are supported by high values of intercept (constant value) with high values of standard error. If we consider the sites where the slopes obtained with S1 are higher than in GHSL-Landsat (e.g. Amsterdam, Milano, New York and Warsaw), we observe much higher values of the intercept for GHSL-Landsat.

Taking into account all the validation experiments and results which show that the ESM was the closest to reference footprints and considering that the reference building footprints are of high quality, it can be concluded that the ESM can be used as a reference dataset for future European urban studies.

Table 3. Comparison of coefficients derived from linear regression for ESM. Standard errors of coefficients are also provided.

City	ESM		
	Coefficients		Std. error
Torino	Intercept	1825.93	770.04
	Slope	0.91	0.01
Milano	Intercept	14318.55	729.74
	Slope	0.91	0.01
Novara	Intercept	7763.31	1801.61
	Slope	1.12	0.03
Warsaw	Intercept	5572.55	347.42
	Slope	2.02	0.01
Oslo	Intercept	3777.99	302.07
	Slope	1.82	0.01
Montpellier	Intercept	1600.39	50.51
	Slope	1.76	0.003
Amsterdam	Intercept	6160.01	250.42
	Slope	1.51	0.01

Table 4. Comparison of the regression coefficients and their standard errors for 13 validated cities and for S1, S2 and Landsat.

City	S1			S2			GHSL-Landsat		
	Coefficients		Std. error	Coefficients		Std. error	Coefficients		Std. error
Torino	Intercept	11072.6	2644.91	Intercept	37056.5	2471.15	Intercept	67700	3805.57
	Slope	1.82	0.04	Slope	1.08	0.03	Slope	1.63	0.05
Milano	Intercept	74991.9	2677.74	Intercept	34808.5	1495.62	Intercept	129655	2896.38
	Slope	1.57	0.04	Slope	0.91	0.02	Slope	1.16	0.04
Novara	Intercept	6847.75	3514.09	Intercept	10596.3	2258.54	Intercept	32878	4092.69
	Slope	1.92	0.06	Slope	1.18	0.04	Slope	2.09	0.07
Warsaw	Intercept	10463.5	928.48	Intercept	14996.8	704.32	Intercept	23727.9	1055.85
	Slope	4.49	0.04	Slope	2.95	0.03	Slope	4.32	0.04
Oslo	Intercept	8725.44	971.15	Intercept	6379.01	592.64	Intercept	13305	1036.33
	Slope	2.96	0.04	Slope	1.9	0.02	Slope	3.67	0.04
Montpellier	Intercept	-860.22	93.41	Intercept	677.31	73.25	Intercept	3752.34	146.05
	Slope	2.68	0.01	Slope	2	0.01	Slope	3.18	0.01
Amsterdam	Intercept	29977.2	741.59	Intercept	17324.2	514.96	Intercept	43656.5	791.59
	Slope	3.03	0.03	Slope	2	0.02	Slope	2.86	0.03
Aizuwakamatsu	Intercept	-1625.7	483.41	Intercept	2278.95	282.17	Intercept	1478.78	480.76
	Slope	3.06	0.03	Slope	2.8	0.02	Slope	3.24	0.03
Dar es Salaam	Intercept	71524.3	1996.03	Intercept	49884	1494.26	Intercept	65979.9	2205.36
	Slope	1.86	0.04	Slope	1.98	0.03	Slope	2.38	0.05
Glenorchy	Intercept	17884.2	2229.99	Intercept	9120.52	1803.05	Intercept	15821.1	2294.59
	Slope	4.26	0.12	Slope	3.64	0.1	Slope	5.29	0.13
New York	Intercept	76377.1	1360.59	Intercept	31317.6	1253.89	Intercept	127422	1512.88
	Slope	2.1	0.02	Slope	1.95	0.02	Slope	1.61	0.03
Surrey	Intercept	6473.77	1334.62	Intercept	6906.48	1177.49	Intercept	28607.1	1590.91
	Slope	1.45	0.04	Slope	2.25	0.04	Slope	3.73	0.05
Washington	Intercept	54003.9	2860.76	Intercept	39225.7	2621.87	Intercept	136900	3555.08
	Slope	2.56	0.06	Slope	2.32	0.06	Slope	1.79	0.08

References

- Ban, Y., Jacob, A., Gamba, P., 2015. Spaceborne SAR data for global urban mapping at 30m resolution using a robust urban extractor. *ISPRS J. Photogramm. Remote Sens.* 103, 28–37. doi:10.1016/j.isprsjprs.2014.08.004
- Chen, Jun, Chen, Jin, Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., Mills, J., 2015. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens., Global Land Cover Mapping and Monitoring* 103, 7–27. doi:10.1016/j.isprsjprs.2014.09.002
- Chrysoulakis, N., Feigenwinter, C., Triantakoustantis, D., Penyeveskiy, I., Tal, A., Parlow, E., Fleishman, G., Düzgün, S., Esch, T., Marconcini, M., 2014. A Conceptual List of Indicators for Urban Planning and Management Based on Earth Observation. *ISPRS Int. J. Geo-Inf.* 3. doi:10.3390/ijgi3030980
- Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* 20, 37–46. doi:10.1177/001316446002000104
- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* 37, 35–46.
- Esch, T., Marconcini, M., Felbier, A., Roth, A., Heldens, W., Huber, M., Schwinger, M., Taubenbock, H., Muller, A., Dech, S., 2013. Urban Footprint Processor - Fully Automated Processing Chain Generating Settlement Masks From Global Data of the TanDEM-X Mission. *IEEE Geosci. Remote Sens. Lett.* 10, 1617–1621. doi:10.1109/LGRS.2013.2272953
- Ferri, S., Halkia, M., Siragusa, A., 2017. The European Settlement Map 2017 Release; Methodology and output of the European Settlement Map (ESM2p5m) (No. JRC105679), JRC Scientific and Technical Report.
- Ferri, S., Syrris, V., Florczyk, A., Scavazzon, M., Halkia, M., Pesaresi, M., 2014. A new map of the European settlements by automatic classification of 2.5m resolution SPOT data. Presented at the Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, pp. 1160–1163.
- Florczyk, A.J., Ferri, S., Syrris, V., Kemper, T., Halkia, M., Soille, P., Pesaresi, M., 2016. A New European Settlement Map From Optical Remotely Sensed Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 9, 1978–1992. doi:10.1109/JSTARS.2015.2485662
- Foody, G.M., 2008. Harshness in image classification accuracy assessment. *Int. J. Remote Sens.* 29, 3137–3158. doi:10.1080/01431160701442120
- Jeni, L.A., Cohn, J.F., De La Torre, F., 2013. Facing Imbalanced Data Recommendations for the Use of Performance Metrics. *Int. Conf. Affect. Comput. Intell. Interact. Workshop Proc. ACII Conf.* 2013, 245–251. doi:10.1109/ACII.2013.47
- Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 159–174.
- Matsuoka, M., Yamazaki, F., 2004. Use of Satellite SAR Intensity Imagery for Detecting Building Areas Damaged Due to Earthquakes. *Earthq. Spectra* 20, 975–994. doi:10.1193/1.1774182

- Pesaresi, Martino, Corbane, C., Julea, A., Florczyk, A., Syrris, V., Soille, P., 2016a. Assessment of the Added-Value of Sentinel-2 for Detecting Built-up Areas. *Remote Sens.* 8, 299. doi:10.3390/rs8040299
- Pesaresi, M., Ehrlich, D., 2009. A methodology to quantify built-up structures from optical VHR imagery, in: Gamba, P., Herold, M. (Eds.), *Global Mapping of Human Settlement Experiences, Datasets, and Prospects*. CRC Press, pp. 27–58.
- Pesaresi, M., Ehrlich, D., Caravaggi, I., Kauffmann, M., Louvrier, C., 2011. Towards Global Automatic Built-Up Area Recognition Using Optical VHR Imagery. *Sel. Top. Appl. Earth Obs. Remote Sens. IEEE J. Of* 4, 923–934.
- Pesaresi, M., Ehrlich, D., Ferri, S., Florczyk, A., Carneiro Freire Sergio, M., Halkia, S., Julea, A., Kemper, T., Soille, P., Syrris, V., 2016a. Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014. Publications Office of the European Union.
- Pesaresi, M., Gerhardinger, A., Kayitakire, F., 2008. A Robust Built-Up Area Presence Index by Anisotropic Rotation-Invariant Textural Measure. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 1, 180–192. doi:10.1109/JSTARS.2008.2002869
- Pesaresi, M., Huadong, G., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., Halkia, M., Kauffmann, M., Kemper, T., Lu, L., Marin-Herrera, M.A., Ouzounis, G.K., Scavazzon, M., Soille, P., Syrris, V., Zanchetta, L., 2013. A Global Human Settlement Layer From Optical HR/VHR RS Data: Concept and First Results. *Sel. Top. Appl. Earth Obs. Remote Sens. IEEE J. Of* 6, 2102–2131.
- Pesaresi, M., Syrris, V., Julea, A., 2016b. A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning. *Remote Sens.* 8, 399. doi:10.3390/rs8050399
- Pesaresi, Martino, Syrris, V., Julea, A., 2016b. Analyzing big remote sensing data via symbolic machine learning, in: *Proceedings of the 2016 Conference on Big Data from Space*. Santa Cruz de Tenerife (Spain), pp. 156–159.
- Pesaresi, M., Vasileios, S., Julea, A., 2016c. Analyzing big remote sensing data via symbolic machine learning., in: *Proceedings of the 2016 Conference on Big Data from Space (BiDS'16)*. Presented at the Big Data from Space (BiDS'16), pp. 156–159. doi:10.2788/854791
- Triantakoustantis, D., Chrysoulakis, N., Sazonova, A., Esch, T., Feigenwinter, C., Düzgün, S., Parlow, E., Marconcini, M., Tal, A., 2015. On-line Evaluation of Earth Observation Derived Indicators for Urban Planning and Management. *Urban Plan. Des. Res.* 3, 17–17.
- Wenkai Li, Qinghua Guo, 2014. A New Accuracy Assessment Method for One-Class Remote Sensing Classification. *IEEE Trans. Geosci. Remote Sens.* 52, 4621–4632. doi:10.1109/TGRS.2013.2283082

List of abbreviations and definitions

ESM – European Settlement Map

GHSL – Global Human Settlement Layer

S1 – Sentinel -1

S2 – Sentinel -2

OSM – Open Street Map

ENDI - Evidence-based Normalized Differential Index

SML – Symbolic Machine Learning

UA – Urban Atlas

CLC – Corine Land Cover

SAR – Synthetic Aperture Radar

DEM – Digital Elevation Model

SSL – Soil sealing layer

GRD – Ground Range Detected

List of figures

Figure 1. Conceptual processing flow for Symbolic Machine Learning. (A) The image to be classified (X) and the reference training layer (Y); (B) the features , (C) the data sequences, (D) the frequency of association; E) the confidence index; and (F) the built-up (figure modified from Pesaresi et al., 2016b).....	4
Figure 2. The Landsat GHSL workflow is broken down into four parts. Cloud removal (A), classification of built-up and water (B), data reduction and single date mosaic production (C) and the multi-temporal fusion (D).....	5
Figure 3. Simplified workflow showing the adaptation of the SML to the classification of Sentinel-1 images at the global level. The input features comprise 18 features derived from dual-polarization Sentinel-1 intensity data and 3 topographic features derived from a global digital elevation model (i.e. SRTM).	6
Figure 4. Simplified workflow showing the adaptation of the SML to the classification of built-up from Sentinel-2 BOA images (excluding the extraction of additional thematic layers e.g. vegetation, water, etc.). The input features comprise reflectance values of the four 10 m bands resampled at 20 m and the reflectance values of the 20m bands. The original 10 bands are used for deriving a textural feature that is used for post-processing the ENDI confidence and refining the detection of built-up.....	7
Figure 5. Selected cities for the validation experiment.	9
Figure 6. Comparison of different cell sizes for cumulative built-up curve analysis in Amsterdam study site: a) 250 m, b) 500 m and c) 1000 m. Reference (black line) is the "reference curve".	11
Figure 7. Performance metrics for European cities. Black lines correspond to the average value of a given measure.	13
Figure 8. Performance metrics for all cities. Averaged values by city are represented with black line.	14
Figure 9. Cumulative built-up curve analysis for European cities. The x axis corresponds to decreasing reference built-up densities calculated for grid cells of 500x500m. Y axis is the cumulative built-up for all layers, including the reference layer.....	15
Figure 10. Results of the cumulative built-up curve analysis for all the study sites.....	16
Figure 11. Estimated built-up area per layer and reference built-up area for all sites ...	17
Figure 12. Scatterplot matrices with correlation coefficients for European study. X axis corresponds to reference built-up extracted from 500x500m grid cells, whilst Y axis corresponds to built-up area obtained from grid cells for specific layers (ESM, S1, S2 and GHSL-Landsat)	19
Figure 13. Scatterplot matrices with correlation coefficient in non-European cities.....	20
Figure 14. Slopes from the linear regression model plotted for each European city and for each product.	21
Figure 15. Slope values plotted for each validated site.	22
Figure 16. Close view of built-up derived from S1, S2 and Landsat in Amsterdam (location: 52.30N, 4.76E).....	23
Figure 17. Close view of built-up derived from S1, S2 and Landsat in Amsterdam (location: 52.41N, 4.77E).....	24
Figure 18. Close view of built-up derived from S1, S2 and Landsat in Amsterdam (location: 52.45N, 4.70E).....	25

Figure 19. Close view of built-up derived from S1, S2 and Landsat with reference footprints in Aizuwakamatsu (location: 37.42N, 139.81E).	26
Figure 20. Close view of built-up derived from S1, S2 and Landsat with reference footprints in Aizuwakamatsu (location: 37.47N, 139.96E).	27
Figure 21. Close view of built-up derived from S1, S2 and Landsat in New York (location: 40.57N, -73.95E).	28
Figure 22. Close view of built-up derived from S2 in Amsterdam (location: 52.23N, 4.74E)	29
Figure 23. Close view of built-up derived from the ESM in Oslo (location: 59.92N, 10.74E).	30
Figure 24. Close view of built-up derived from S1, S2 and Landsat in Surrey study site (location: 49.12N, -122.91E).	31

List of tables

Table 1. Characteristics of the built-up products under validation	8
Table 2. Study sites with corresponding spatial extent, source of reference dataset and the year of reference dataset.....	9
Table 3. Comparison of coefficients derived from linear regression for ESM. Standard errors of coefficients are also provided.....	34
Table 4. Comparison of the regression coefficients and their standard errors for 13 validated cities and for S1, S2 and Landsat.	35

Europe Direct is a service to help you find answers to your questions about the European Union
Free phone number (*): 00 800 6 7 8 9 10 11
(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed.

A great deal of additional information on the European Union is available on the Internet.
It can be accessed through the Europa server <http://europa.eu>

How to obtain EU publications

Our publications are available from EU Bookshop (<http://bookshop.europa.eu>),
where you can place an order with the sales agent of your choice.

The Publications Office has a worldwide network of sales agents.
You can obtain their contact details by sending a fax to (352) 29 29-42758.

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.



EU Science Hub

ec.europa.eu/jrc



@EU_ScienceHub



EU Science Hub - Joint Research Centre



Joint Research Centre



EU Science Hub

